Separating Hash Families

Dissertation
zur Erlangung des Grades eines Doktors
der Naturwissenschaften (Dr. rer. nat.)

dem Fachbereich Mathematik
der Universität Duisburg-Essen

von
Marjan Bazrafshan
aus dem Iran

Datum der mündlichen Prüfung: 10. Juni 2011
Vorsitzender: Prof. Dr. Wolfgang Lempken
Gutachter:
Prof. Dr. Trung van Tran
Prof. Dr. Alfred Wassermann, Universität Bayreuth

Zusammenfassung

In der vorliegenden Dissertation wird angestrebt, offene Probleme im Zusammenhang mit sogenannten "separating hash families" zu diskutieren und zu lösen.

Separating hash families (SHF) sind interessante kombinatorische Strukturen, die verschiedene bekannte Objekte als Spezialfälle einschließen, wie z.B. *perfect hash families* (PHF), *frameproof codes*, *secure frameproof codes* und *codes with identifiable parent property*. Ferner finden SHFs zahlreiche kryptographische Anwendungen, z.B. in *key distribution patterns, broadcast encryption, secret sharing schemes, visual cryptography* und in den Codes für den Urheberrechtsschutz.

Eine separating hash family, $\mathsf{SHF}(N; n, m, \{w_1, w_2, ..., w_t\})$ mit dem Typ $\{w_1, \ldots, w_t\}$, $t \geq 2$, lässt sich als eine $N \times n$ Matrix $\mathcal{A}$ mit Einträgen aus einer Menge von m Symbolen beschreiben, so dass es für alle disjunkten Mengen $\mathsf{C}_1, \ldots, \mathsf{C}_t$ von Spalten von $\mathcal{A}$ mit $|\mathsf{C}_i| = w_i$, mindestens eine Zeile r von $\mathcal{A}$ derart gibt, dass für alle $i \neq j$ die Mengen $\{\mathcal{A}(r, x) : x \in \mathsf{C}_i\}$ und $\{\mathcal{A}(r, x) : x \in \mathsf{C}_j\}$ disjunkt sind.

Das erste Hauptproblem bei der Untersuchung von SHFs ist die Bestimmung guter Schranken für die Maximalzahl der Spalten n, wenn die anderen Parameter N, m und $\{w_1, \ldots, w_t\}$ gegeben sind. Das zweite Hauptproblem ist die Konstruktion von "guten" SHFs. Diese beiden Probleme sind im Allgemeinen schwierig zu lösen. Die aus der Literatur bekannten Schranken für SHFs sind noch weit entfernt davon, scharf zu sein.

In dieser Dissertation konzentrieren wir uns auf die Herleitung oberer Schranken für die Anzahl der Spalten einer SHF. Zuerst werden spezifische Typen von SHFs untersucht und einige ihrer Eigenschaften bewiesen. Basierend darauf erzielen wir neue obere Schranken für die maximale Anzahl der Spalten bzw. untere Schranken für die minimale Anzahl der Zeilen einer SHF. Für bestimmte Parameter geben wir Konstruktionen von SHFs, so dass die erzielten Schranken mit Gleichheit erfüllt sind. Damit sind die Schranken im oberen Fall optimal. Für den Typ $\{2, 2\}$ präsentieren wir eine generelle Konstruktion, die für große m asymptotisch optimal ist.

Anschließend untersuchen wir generelle SHFs und stellen drei neue obere Schranken vor, die schärfer als alle bisher bekannten Schranken sind. Um die erste Schranke herzuleiten, wird zunächst gezeigt, wie eine neue SHF von einer existierenden SHF

bei einzeiliger Entfernung hergeleitet werden kann. Darauf basierend können wir den Beweis der ersten Schranke führen. Die zweite Schranke wird durch Modifizierung einer bereits existierenden Methode erzielt. Unsere dritte Schranke ist eine Verallgemeinerung von einer bekannten Schranke. Ferner untersuchen wir weitere Fälle, für die schärfere Schranken für SHFs nachgewiesen werden können. Schließlich liefern wir mehrere neue rekursive Konstruktionen für SHFs.

Acknowledgment

I give my deepest thanks to my supervisor Professor Tran van Trung for supporting me so patiently to complete this research. I appreciate all his helpful advice and guidance on mathematical research and also all the enjoyable discussions on many different topics. He has always been available and spent so much time helping me write this dissertation.

I wish further to appreciate Professor Han Vinck for the friendly atmosphere he has created in the Digital Communications Group. I am specially thankful to him for giving me the chance to participate in doing some projects on biometrics which made me familiar with this interesting field and also helped me financially.

I am also thankful to my dear colleagues Anil, Julia, Pavol, Birgit, Anahit and Balakirsky and my lovely friends in the IEM, Lukas, Hamutal and Lior for their friendship and accompanying.

I owe many thanks to my family who have never let me alone in life. My parents encouraged me to continue graduate studies and always supported me in this way.

Last but not least I would like to thank my husband Navid for his sincere love which makes life much more enjoyable.

Contents

List of Tables

Chapter 1

Introduction

In this thesis, we focus on *separating hash families*. Our motivation for studying this combinatorial object is the generality of the concept. Separating hash family was defined in [44] for the first time. However, different various objects which have been studied since a long time before then, have been found to be special cases of separating hash families. Hence, investigating separating hash families results in interesting generalization of previously known facts.

Cryptography is one of the areas in which various kinds of separating hash families are used. They are applied in constructing certain kinds of fingerprinting codes. We introduce this line of research which was initiated by Boneh and Shaw [19] and Hollmann et al [31], very briefly in the following.

One serious problem, which information providers encounter, is the unauthorized redistribution of copyrighted materials. With the increasing application of digital data, such as documents, images, movies, music, computer software, etc., the problem of protecting them against illegal copying becomes more important. An old cryptographic technique used to prevent users from piracy is *Fingerprinting*. To fingerprint a digital product, a distributor assigns to each copy of the product some unique codeword and sends the product to the users *marked* with that codeword. He also saves a database of sold copies and their corresponding fingerprints for himself. Later, if the distributor discovers an illegally sold copy of a product, he can easily trace it back to the user who owns the original copy by comparing its fingerprint to the database.

An offending user may try different types of attacks to distribute illegal copies anonymously. One strong attack occurs when several users collude and compare the fingerprints assigned to their copies. They detect the locations where their fingerprints differ. These locations are called *marks*. They can then produce an illegal copy assigned with a new fingerprint different from all their codewords. This new codeword is obtained by changing the elements on the locations belonging to the subset of marks they have found. In this way, the pirates cause some problems for the distributor who wants to trace a piracy back.

One problem is that, by this attack the pirates may *frame* a user not in the coalition. Hence, when the distributor detects a piracy, by tracing the codeword back, an innocent user would be recognized as a pirate. In order to overcome this problem, Boneh and Shaw [19] defined the concept of *frameproof* codes. A *w-frameproof* code has the property that no coalition of at most w users can frame a user not in the coalition.

The second possible problem is that two disjoint coalitions may be able to produce a common descendant. Therefore, tracing the fingerprint of an illegal copy back, may result in two different sets of users and makes it impossible to recognize the pirates. *Secure frameproof* codes [44] protect against disjoint coalitions to be able to make the same fingerprint.

A stronger version of protection is provided by *identifiable parent property* (IPP) codes [31]. IPP codes have the property that if there exist several coalitions which are able to produce a common descendant, then the coalitions must have a codeword in common (the common codeword is the *identifying parent*), i.e. by tracing the common descendant back, one guilty user is certainly recognized.

The codes introduced above, are all special cases of separating hash families.

Another important special case of separating hash families is the class of *perfect hash families* which are useful tools for cryptographers. A perfect hash family is a family of functions sharing a domain and codomain with the property that there exists always one function having different outputs for a fixed number of single inputs.

Perfect hash families have a wide range of applications. Mehlhorn [37, 38] introduced this concept in compiler design to get lower bounds on the size of a computer program. In cryptography, they are used in threshold secret sharing [11, 16], broadcast encryption [29], shared symmetric key primitives [33, 34], private information retrieval [10], construction of fingerprinting codes [42] and key distribution patterns [43, 44]. Moreover, they have been applied to circuit complexity problems [39] and construction of deterministic analogues of probabilistic algorithms [2]. In [44], perfect hash families are used to improve explicit constructions of secure frameproof codes, key distribution patterns, group testing algorithms, cover free families and separating systems. They have also applications to operating system, language translation system, hypertext, hypermedia, file managers and information retrieval system [25].

In the definition of perfect hash families explained above, instead of a fixed number of single inputs we can consider a number of disjoint sets of inputs. In this way we generalize the concept of perfect hash families to separating hash families. Specifically, a separating hash family is a family of functions having the same domain and range with the property that there exists at least one function having different output sets for every fixed number of disjoint sets of inputs with specified cardinalities.

We can represent each separating hash family with a matrix as explained in the following. Assume that $\mathcal{F}$ is a family of N functions defined from a set X to a set Y with cardinalities $|X| = n$ and $|Y| = m$. The matrix representation $\mathcal{A}$ of $\mathcal{F}$ is an $N \times n$ matrix in which each row corresponds to a function $f \in \mathcal{F}$ and each column corresponds to a member $a \in X$. The entry in row f and column a is $f(a)$. The separating property of the hash family $\mathcal{F}$ implies that if $\mathsf{C}_1, \mathsf{C}_2, \ldots, \mathsf{C}_t$ are disjoint sets of columns of $\mathcal{A}$ with $|\mathsf{C}_1| = w_1, |\mathsf{C}_2| = w_2, \ldots, |\mathsf{C}_t| = w_t$, then $f(\mathsf{C}_1), f(\mathsf{C}_2), \ldots, f(\mathsf{C}_t)$ are disjoint sets.

The interesting problem in studying separating hash families is to find a relationship between N and n when m and $\{w_1, w_2 \ldots, w_t\}$ are known. Given N, we are interested in the maximum possible number of n. Equivalently, when n is fixed, we want to obtain a minimum value on N.

In this thesis we focus mostly on the problem explained in the previous paragraph. We consider separating hash families with different parameters and try to obtain new upper bounds on the number of its columns or lower bounds on the number of rows. In some cases we improve previously known bounds and obtain stronger bounds or extend existing bounds for special cases to more general parameters. We also provide new upper bounds on the number of columns of the array corresponding to a general separating hash family of arbitrary parameters. For some types of separating hash families, we present constructions to show that the obtained bounds are tight (i.e. the bound is achievable) which implies that the construction is optimal. In addition, we investigate some constructions of separating hash families. We provide both recursive and direct constructions for separating hash families.

Thesis overview

This thesis is organized as follows.

Chapter 2 is a brief survey on separating hash families. In 2.1 we define separating hash families and some other combinatorial objects which are special cases of separating hash families. We give basic properties of separating hash families which are useful in studying them in 2.2 and then in 2.3 present some examples of separating hash families.

Chapter 3 is a summary of the known upper bounds on the number of columns of separating hash families. In this chapter we present the strong previously known upper bounds for different types of separating hash families. These results will be later compared with our new results to show that they are improved.

In Chapter 4 we investigate some separating hash families having small types. We consider the types $\{3, 2\}$ in Section 4.1, $\{3, 3\}$ in 4.2 and $\{1, 2\}$ in 4.3. For each

type we first prove some properties of these separating hash families which help us to obtain new bounds on the number of columns for the considered types. Moreover, for the type $\{1, 2\}$ an optimal construction is provided.

We present new results on separating hash families of type $\{2, 2\}$ in Chapter 5. In order to prove a new bound in this case, we first consider separating hash families with three rows and prove some properties of such separating hash families in 5.1. In 5.2, we present some optimal separating hash families with three rows and small number of symbols. Then, in 5.3 we prove our new bound. Finally, we give some constructions for separating hash families of type $\{2, 2\}$. The results in Chapters 4 and 5 are presented in [8].

Chapter 6 gives new bounds on the number of columns of separating hash families of general type. Our new bounds presented in this chapter, improve all the previously known bounds on general separating hash families.

In Chapter 7, we present some ways to improve the general bounds in special cases. The results of this chapter show that the general bounds are not strong enough in all cases.

Chapter 8 presents some constructions for separating hash families. In this chapter we give recursive constructions for constructing new separating hash families from the existing ones. Our first construction uses a 2-separating hash family and produces a new 2-separating hash family with increased number of rows and columns. The second construction is a generalization of a recursive construction for perfect hash families.

Finally, in Chapter 9, some ideas for the possible future works are presented.

Chapter 2

Preliminaries

The concept of *separating hash families* was first defined in [44]. Later, Cohen et al. [22] proved a sufficient condition on an arbitrary code with large minimum distance to be a separating hash family of a specified type. Separating hash families with small parameters were thoroughly studied by Stinson et al. in [45]. Stinson and Zaverucha [47] generalized some of the results of [45] to a larger class of separating hash families. In [17], this study was extended to separating hash families of arbitrary type.

The aim of this chapter is to introduce separating hash families in general. In 2.1, we define the notions of separating hash families and some other combinatorial structures and investigate the connection between them. We will see how separating hash families include other useful objects as special cases. In 2.2 some properties of separating hash families which are useful in studying them are presented. Finally, in 2.3, we give some constructions of separating hash families. Consequently, we show some parameters for which a separating hash family exists.

2.1 Definitions

In this section some basic definitions which are necessary in this thesis are presented. We first define the concept of a hash family which is a very general object. Then we introduce the special cases perfect hash families and separating hash families. We also give the definition of different types of fingerprinting codes. Finally, we present the relationship between separating hash families and those constructions which are special cases of this family of functions. This connection gives motivation for studying separating hash families.

Roughly speaking, a *hash family* is a collection of functions sharing the same domain and range. A function is said to *separate* two or more different elements of the domain, if it assigns different values to them. A hash family which has at least one separating function for every fixed number of elements of its domain, is called a *perfect hash family*. This concept of separating can be naturally generalized to two or more input

subsets, i.e. a function separates some disjoint subsets of its domain, if their images are disjoint sets. A hash family is called *separating* if for every fixed number of disjoint subsets of its domain with specified cardinalities, there exists at least one function which separates them. In the following we present the precise definitions of these concepts.

2.1.1 Hash families

We begin with the definition of hash families.

Definition 2.1.1. Let X and Y be two sets with $|X| = n$ and $|Y| = m$ and $\mathcal{F}$ be a set of functions from X to Y with $|\mathcal{F}| = N$. We call $\mathcal{F}$ an $(N; n, m)$-*hash family* and denote it by $(N; n, m)$-HF.

For simplicity, we usually assume $X = \{1, 2, \ldots, n\}$. In general, an $(N; n, m)$-HF $\mathcal{F}$ can be depicted as an $N \times n$ array $\mathcal{A}$ having entries on a set of m symbols. The rows of $\mathcal{A}$ correspond to the functions in the family and the columns are representatives of the elements in the domain set X. The entry in row corresponding to the function f and column a of the array is $f(a)$. $\mathcal{A}$ is called the *matrix representation* of $\mathcal{F}$. We refer to the entries of $\mathcal{A}$ (i.e. elements of Y) as *symbols*.

A hash family can also be considered as a *code*. We first define the notion of a code and give the notations that we use, and then correspond a code to every hash family.

Definition 2.1.2. Let $\mathcal{Q}$ be a set of *alphabets* of size m. A subset $\mathcal{C} \subseteq \mathcal{Q}^N$ with $|\mathcal{C}| = n$ is a *code* of length N on $\mathcal{Q}$ and the elements of $\mathcal{C}$ are called *codewords*. A code $\mathcal{C}$ with above parameters is denoted as an $(N; n, m)$-code. For every two codewords $c_1, c_2 \in \mathcal{C}$, the *distance* between c_1 and c_2 (also called the *Hamming distance* of c_1 and c_2) is the number of positions in which c_1 and c_2 differ and is denoted by $\mathrm{dist}(c_1, c_2)$. The *minimum distance* of a code is defined to be the smallest distance between two codewords, i.e.

$$\mathrm{dist}(\mathcal{C}) = \min\{\mathrm{dist}(c_1, c_2) : c_1, c_2 \in \mathcal{C}, c_1 \neq c_2\}.$$

An $(N; n, m)$-code with minimum distance d is denoted as $(N; n, m, d)$-code.

Let $\mathcal{F}$ be an $(N; n, m)$-HF with domain X and range Y and $\mathcal{A}$ be the matrix representation of $\mathcal{F}$. We can consider each column of $\mathcal{A}$ as a codeword and view $\mathcal{A}$ as an $(N; n, m)$-code over the alphabet Y.

2.1.2 Perfect hash families

Perfect hash families have been investigated by many researchers for a long time. For some results in this area see [4, 9, 12, 18, 26, 46, 49, 51].

To define the concept of perfect hash families, we need the definition of a *separating* function.

Definition 2.1.3. Let $f : X \to Y$ be a function and $T \subseteq X$. f is said to *separate* T if f is injective on T. We also say that f is *separating* on T.

Here is the definition of perfect hash families.

Definition 2.1.4. Let t be a positive integer. An $(N; n, m)$-HF $\mathcal{F}$ on the domain set X is an $(N; n, m, t)$-*perfect hash family* if for all $T \subseteq X$ with $|T| = t$ there exists at least one function $f \in \mathcal{F}$ such that f separates T. t is called the *strength* of $\mathcal{F}$.

A perfect hash family (PHF) with above parameters is denoted by $\mathsf{PHF}(N; n, m, t)$.

It is easily observed that the matrix representation of a $\mathsf{PHF}(N; n, m, t)$ is an $N \times n$ array with entries from a set of m symbols such that every $N \times t$ subarray contains at least one row having distinct symbols.

If $t = 1$, then every array with even one row has the property of a perfect hash family. It means that there exists a $\mathsf{PHF}(1; n, m, 1)$ for every positive integer n and m. In order to avoid the trivialities, we usually assume that $t > 1$ in a $\mathsf{PHF}(N; n, m, t)$.

We present an example of a $\mathsf{PHF}(6; 8, 4, 4)$ [35]. This example was found by computer (Tran van Trung). Martirosyan showed that for any $\mathsf{PHF}(6; n, 4, 4)$ we have $n \leq 8$. It means that the following PHF has the maximum possible number of columns.

Example 2.1.5. [35] Let $X = \{1, 2, \ldots, 8\}$, $Y = \{1, 2, 3, 4\}$ and $\mathcal{F} = \{f_1, f_2, f_3, f_4, f_5, f_6\}$ where for $i = 1, \ldots, 6$, f_i is defined below:

$$
\begin{array}{llll}
f_1(1) = f_1(3) = 1 & , \; f_1(2) = f_1(4) = 2 & , \; f_1(5) = f_1(7) = 3 & , \; f_1(6) = f_1(8) = 4 \\
f_2(1) = f_2(2) = 1 & , \; f_2(3) = f_2(4) = 2 & , \; f_2(5) = f_2(6) = 3 & , \; f_2(7) = f_2(8) = 4 \\
f_3(1) = f_3(6) = 1 & , \; f_3(4) = f_3(7) = 2 & , \; f_3(3) = f_3(5) = 3 & , \; f_3(2) = f_3(8) = 4 \\
f_4(2) = f_4(7) = 1 & , \; f_4(1) = f_4(5) = 2 & , \; f_4(3) = f_4(8) = 3 & , \; f_4(4) = f_4(6) = 4 \\
f_5(4) = f_5(5) = 1 & , \; f_5(3) = f_5(7) = 2 & , \; f_5(2) = f_5(6) = 3 & , \; f_5(1) = f_5(8) = 4 \\
f_6(1) = f_6(7) = 1 & , \; f_6(2) = f_6(5) = 2 & , \; f_6(3) = f_6(6) = 3 & , \; f_6(4) = f_6(8) = 4
\end{array}
$$

The matrix representation of $\mathcal{F}$ is as follows:

	1	2	3	4	5	6	7	8
f_1	1	2	1	2	3	4	3	4
f_2	1	1	2	2	3	3	4	4
f_3	1	4	3	2	3	1	2	4
f_4	2	1	3	4	2	4	1	3
f_5	4	3	2	1	1	3	2	4
f_6	1	2	3	4	2	3	1	4

The $(6; 8, 4)$-code $\mathcal{C}$ corresponding to $\mathcal{F}$ is:

$$
\begin{aligned}
\mathcal{C} = \{&(1,1,1,2,4,1), (2,1,4,1,3,2), (1,2,3,3,2,3), (2,2,2,4,1,4), (3,3,3,2,1,2), \\
&(4,3,1,4,3,3), (3,4,2,1,2,1), (4,4,4,3,4,4)\}.
\end{aligned}
$$

2.1.3 Separating hash families

In [44], separating hash family was defined as a hash family that separates two input sets of specific cardinalities. In [45] Stinson et al. generalized this notion to every arbitrary number of sets.

Definition 2.1.6. Assume that $\mathcal{F}$ is an $(N; n, m)$-HF. $\mathcal{F}$ is a *separating hash family* (SHF) of parameters $(N; n, m, \{w_1, w_2, \ldots, w_t\})$ if it satisfies the following property:

For any disjoint subsets $\mathsf{C}_1, \mathsf{C}_2, \ldots, \mathsf{C}_t \subseteq X$ with $|\mathsf{C}_1| = w_1$, $|\mathsf{C}_2| = w_2$, $\ldots$, $|\mathsf{C}_t| = w_t$, there exists at least one $f \in \mathcal{F}$ such that

$$\{f(a) : \ a \in \mathsf{C}_i\} \cap \{f(a) : \ a \in \mathsf{C}_j\} = \emptyset,$$

for any $i \neq j$. In other words, f *separates* the sets C_i for $i = 1, \ldots, t$.

We use the notation $\mathsf{SHF}(N; n, m, \{w_1, w_2, \ldots, w_t\})$ to denote a separating hash family with above parameters.

The multiset $\{w_1, w_2, \ldots, w_t\}$ is called the *type* of the separating hash family.

It is easy to observe that a $\mathsf{PHF}(N; n, m, t)$ is an $\mathsf{SHF}(N; n, m, \{1, \ldots, 1\})$.

If $\mathcal{F}$ is an $\mathsf{SHF}(N; n, m, \{w_1, w_2, \ldots, w_t\})$ and $\mathcal{A}$ is the matrix representation of $\mathcal{F}$, then $\mathcal{A}$ satisfies the following property:

For any disjoint sets of columns C_1, C_2, $\ldots$, C_t, with $|\mathsf{C}_i| = w_i$ for $i = 1, \ldots, t$, there exists a row r of $\mathcal{A}$ such that

$$\{\mathcal{A}(r, a) : a \in \mathsf{C}_i\} \cap \{\mathcal{A}(r, a) : a \in \mathsf{C}_j\} = \emptyset$$

for all $i \neq j$. In this case, the row r separates the sets C_i for $i = 1, \ldots, t$.

Now we give an example of an SHF.

Example 2.1.7. Here is an example of an $\mathsf{SHF}(3; 16, 8, \{2, 2\})$ represented by its corresponding matrix.

0	0	1	1	2	2	3	3	4	4	5	5	6	6	7	7
0	1	0	1	2	3	2	3	4	5	4	5	6	7	6	7
0	1	1	0	2	3	3	2	4	5	5	4	6	7	7	6

The interesting general question in studying separating hash families is to find a relationship between n and N for a specified type $\{w_1, \ldots, w_t\}$ and alphabet size m. The aim is to obtain the maximum possible n for a given N or equivalently to derive a minimum value on N when n is known.

To identify a separating hash family with maximum value of n we give the following definition.

Definition 2.1.8. Assume that the type $\{w_1, \ldots, w_t\}$ and the alphabet size m are given. An *optimal* separating hash family with parameters $(N; n, m, \{w_1, \ldots, w_t\})$, is a separating hash family with an $N \times n$ matrix representation which has the maximum possible number of columns n when N is given. In other words, for any $\mathsf{SHF}(N; n_1, m, \{w_1, \ldots, w_t\})$ we have $n_1 \leq n$.

To find good upper bounds on n in an $\mathsf{SHF}(N; n, m, \{w_1, \ldots, w_t\})$ is an important subject of this thesis. One of the general methods used to achieve this goal is to show that a particular choice of n means that the matrix representation $\mathcal{A}$ always contains a submatrix which is impossible in an SHF with given parameters. Such a submatrix is referred to as a *forbidden configuration*.

For example, an $\mathsf{SHF}(3; n, m, \{2, 2\})$ cannot contain a submatrix isomorphic to the matrix shown below: (Two matrices are *isomorphic* if one is obtained from the other by a permutation on rows and/or columns and/or symbols.)

a	a	$*$	$*$
b	b	$*$	$*$
$*$	$*$	c	c

in which $*$ denotes an arbitrary symbol and a, b and c are not necessarily distinct. It is easily observed that the first and third columns of the above matrix are not separated from the second and fourth. Hence it is not separating of type $\{2, 2\}$.

Moreover, in our argumentations to obtain a forbidden configuration in an SHF, we use the concept of a *unique* or a *repeating* element which is defined in the following.

Definition 2.1.9. Assume that $\mathcal{F}$ is an $\mathsf{SHF}(N; n, m, \{w_1, \ldots, w_2\})$ and $\mathcal{A}$ is the matrix representation of $\mathcal{F}$. A *unique element* in a row or a column of $\mathcal{A}$ is an element which appears only once in that row or column. An element which appears more than one time in some row or column is called a *repeating* element.

2.1.4 Fingerprinting codes

One of the attractive aspects of separating hash families is their application in fingerprinting digital data. In fact, separating hash families of some specific types are known as codes which provide certain forms of traceability. Hence, they are used in protecting digital data from illegal copying. We introduce three classes of codes with different secure properties namely frameproof codes, secure frameproof codes and codes with identifiable parent property. The concept of frameproof codes was first introduced by Boneh and Shaw [19]. Secure frameproof codes, which are a stronger form of frameproof codes, were introduced in [44]. In [31] the concept of codes with

identifiable parent property is introduced. To define these concepts, we need the following information:

Let $\mathcal{C}$ be an $(N; n, m)$-code and each codeword of $\mathcal{C}$ be of the form $c = (c_1, \ldots, c_N)$, where $c_i \in \mathcal{Q}$ for $i = 1, \ldots, N$.

Definition 2.1.10. Assume that $\mathcal{C}_0$ is an arbitrary subset of $\mathcal{C}$. The set of *descendants* of $\mathcal{C}_0$, denoted by $\text{desc}(\mathcal{C}_0)$, is defined as follows:

$$\text{desc } \mathcal{C}_0 = \{x = (x_1, \ldots, x_N) \in \mathcal{Q}^N \ : \ x_i \in \{c_i : c = (c_1, \ldots, c_N) \in \mathcal{C}_0\}, 1 \leq i \leq N\}.$$

In other words, the set $\text{desc}(\mathcal{C}_0)$ consists of all the N-tuples that can be produced by a coalition of the codewords in $\mathcal{C}_0$.

For a positive integer w and a code $\mathcal{C}$, the *w-descendant code of $\mathcal{C}$*, denoted by $\text{desc}_w(\mathcal{C})$, is defined as follows:

$$\text{desc}_w(\mathcal{C}) = \bigcup_{\mathcal{C}_0 \subseteq \mathcal{C}, |\mathcal{C}_0| \leq w} \text{desc}(\mathcal{C}_0)$$

The elements of the set $\text{desc}_w(\mathcal{C})$ are all the N-tuples that could be produced by some coalition of size at most w.

We now define the following concepts:

Definition 2.1.11. Let $\mathcal{C}$ be an $(N; n, m)$-code and $w \geq 2$ be an integer.

- $\mathcal{C}$ is a *w-frameproof code* (w-FP code) provided that for any subset $\mathcal{C}_0 \subseteq \mathcal{C}$ of cardinality at most w, if $x \in \text{desc}(\mathcal{C}_0) \cap \mathcal{C}$ then $x \in \mathcal{C}_0$. An $(N; n, m)$-code which is a w-FP code is denoted by $(N; n, m, w) - FP$. In fact a code is w-FP if no coalition of size at most w can frame another user not in the coalition by producing the codeword held by that user. For results on frameproof codes, see [13, 19, 23, 27, 28, 32, 40–42, 44, 52].

- $\mathcal{C}$ is a *w-secure frameproof code* (w-SFP code) provided that for any two subsets $\mathcal{C}_0$ and $\mathcal{C}_1$ of $\mathcal{C}$ of cardinality at most w, $\text{desc}(\mathcal{C}_0) \cap \text{desc}(\mathcal{C}_1) \neq \emptyset$ implies that $\mathcal{C}_0 \cap \mathcal{C}_1 \neq \emptyset$. To denote an $(N; n, m)$-code which is a w-SFP code we use the notation $(N; n, m, w) - SFP$. In other words, a code is w-SFP if no coalition of size at most w can frame a disjoint coalition of size at most w by producing an N-tuple that could have been produced by the second coalition (see [21, 23, 44]).

- $\mathcal{C}$ is a code with *w-identifiable parent property* (w-IPP code) if the following condition holds:

  Let $\{\mathcal{C}_1, \ldots, \mathcal{C}_t\}$ be a family of subsets of $\mathcal{C}$ where each $\mathcal{C}_i$, $i = 1, \ldots, t$, is of cardinality at most w. Then

$$\bigcap_{1\leq i\leq t} \mathrm{desc}(\mathcal{C}_i) \neq \emptyset \text{ implies } \bigcap_{1\leq i\leq t} \mathcal{C}_i \neq \emptyset.$$

We use the notation $(N; n, m, w) - IPP$ to denote an $(N; n, m)$-code which has w-IPP. A code has the w-IPP if no coalition of size at most w can produce an N-tuple that cannot be traced back to at least one member of the coalition. w-IPP codes have been studied in [1, 3, 6, 14, 40, 42, 50].

2.1.5 Special cases of SHFs

Now we can present the special cases of separating hash families as follows:

- A $\mathsf{PHF}(N; n, m, t)$ is equivalent to an $\mathsf{SHF}(N; n, m, \{w_1, w_2, \ldots, w_t\})$ where $w_i = 1$ for $i = 1, \ldots, t$.
- An $(N; n, m, w) - FP$ code is equivalent to an $\mathsf{SHF}(N; n, m, \{1, w\})$.
- An $(N; n, m, w) - SFP$ code is equivalent to an $\mathsf{SHF}(N; n, m, \{w, w\})$ where $n \geq 2w$.
- An $(N; n, m, 2) - IPP$ code is equivalent to an $\mathsf{SHF}(N; n, m, \{1, 1, 1\} + \{2, 2\})$.
- An $(N; n, m, w) - IPP$ code is an $\mathsf{SHF}(N; n, m, \{\overbrace{1, \ldots, 1}^{w+1}\} + \{w, w\})$ where $n \geq 2w$ (The converse is proved only for $w = 2$).
- A *Strong* SHF is an SHF of type $\{w_1, w_2, \ldots, w_t\}$ where $w_i = 1$ for all but one i [40] (The converse does not in general hold).

It is observed that SHFs are a general class of functions which include various useful combinatorial objects as special cases. Hence, studying SHFs results in more general facts about these objects.

2.2 Basic properties of SHFs

In this section, we summarize some useful properties of separating hash families. We remark that in this thesis, we identify a separating hash family $\mathcal{F}$ with its matrix representation $\mathcal{A}$ and say sometimes that $\mathcal{F}$ has N rows, n columns, m symbols and type $\{w_1, \ldots, w_t\}$ or $\mathcal{A}$ is an SHF.

First we investigate the parameters N, n, m and $\{w_1, \ldots, w_t\}$ and the relationship between these numbers.

If $n \leq m$, then we can always construct columns which have different symbols in each row, i.e. for every arbitrary N there exists an $\mathsf{SHF}(N; n, m, \{w_1, \ldots, w_t\})$ with $\sum_{i=1}^{t} w_i \leq n$ when $n \leq m$. In particular,

$$\exists\ \mathsf{SHF}(1; n, m, \{w_1, \ldots, w_t\}) \text{ when } n \leq m \text{ and } w_1 + \cdots + w_t \leq n.$$

Hence, in order to avoid trivialities, we always assume that $n > m$. In addition, when $n > m$, then for each row there are at least two columns that have the same symbol. It means that at least two columns are not separated in each row. Hence, there is no separating hash family with $N = 1$ when $n > m$. This result is summarized in the following remark.

Remark 2.2.1. *In an* $\mathsf{SHF}(N; n, m, \{w_1, w_2, \ldots, w_t\})$*, we always assume that* $n > m$ *and consequently* $N \geq 2$*.*

The following two lemmas state obvious necessary conditions for the existence of separating hash families.

Lemma 2.2.2. *In any* $\mathsf{SHF}(N; n, m, \{w_1, w_2, \ldots, w_t\})$*,* $m \geq t$*.*

Lemma 2.2.3. *In any* $\mathsf{SHF}(N; n, m, \{w_1, w_2, \ldots, w_t\})$*,* $\sum_{i=1}^{t} w_i \leq n$*.*

The following lemma states that the rows and columns of a separating hash family can be permuted while the separating property is reserved.

Lemma 2.2.4. *If* $\mathcal{A}$ *is the matrix representation of an* $\mathsf{SHF}(N; n, m, \{w_1, w_2, \ldots, w_t\})$*, and if* $\mathcal{A}'$ *is a matrix obtained from* $\mathcal{A}$ *by permuting the rows and/or columns of* $\mathcal{A}$*, then* $\mathcal{A}'$ *is also an* $\mathsf{SHF}(N; n, m, \{w_1, w_2, \ldots, w_t\})$*.*

Lemma 2.2.4 is the motivation for the following definition:

Definition 2.2.5. Two separating hash families are said to be *isomorphic* if the matrix representation of one is obtained from another by a permutation of rows and/or columns and/or elements.

The following two theorems which show the relationship between separating hash families of different types, follow easily from the definition of separating hash families and are quite useful.

Theorem 2.2.6. *If* $\mathcal{F}$ *is an* $\mathsf{SHF}(N; n, m, \{w_1, w_2, \ldots, w_t\})$*, then* $\mathcal{F}$ *is also an* SHF *of parameters* $(N; n, m, \{w_1', w_2, \ldots, w_t\})$ *for* $w_1' \leq w_1$*.*

Theorem 2.2.7. *If* $\mathcal{F}$ *is an* $\mathsf{SHF}(N; n, m, \{w_1, w_2, \ldots, w_t\})$ *and* $w_1' = w_1 + w_2$*, then* $\mathcal{F}$ *is also an* $\mathsf{SHF}(N; n, m, \{w_1', w_3, \ldots, w_t\})$*.*

The converse of Theorems 2.2.6 and 2.2.7 do not always hold. i.e. an $\mathsf{SHF}(N; n, m, \{w_1', w_2, \ldots, w_t\})$ is not necessarily an $\mathsf{SHF}(N; n, m, \{w_1, w_2, \ldots, w_t\})$ when $w_1' \leq w_1$. In addition for $w_1' = w_1 + w_2$, there exist many examples which show that an $\mathsf{SHF}(N; n, m, \{w_1', w_3, \ldots, w_t\})$ is not an $\mathsf{SHF}(N; n, m, \{w_1, w_2, \ldots, w_t\})$. The example below is a small instance.

Example 2.2.8. The array shown below is an $\mathsf{SHF}(3; 4, 2, \{2, 2\})$.

$$\mathcal{A} = \begin{array}{|c|c|c|c|}\hline 1 & 1 & 0 & 0 \\ \hline 1 & 0 & 1 & 0 \\ \hline 1 & 0 & 0 & 1 \\ \hline \end{array}$$

Let $w_1 = w_2 = 2$, $w'_1 = w'_2 = 1$ and $w'_3 = 2$. Hence $w_1 = w'_1 + w'_2$. It is easy to observe that this array is not separating of type $\{1, 1, 2\}$ which shows that the converse of Theorem 2.2.7 is not true.

In addition, according to Theorem 2.2.6, $\mathcal{A}$ is an $\mathsf{SHF}(3; 4, 2, \{1, 2\})$. Now let $w_1 = 1$, $w_2 = 2$ and $w'_2 = 3 > w_2$. Observing that $\mathcal{A}$ is not an $\mathsf{SHF}(3; 4, 2, \{1, 3\})$ shows that the converse of Theorem 2.2.6 does not hold.

Theorem 2.2.9 gives a simple direct construction of separating hash families from error correcting codes. Stinson et al. [46] proved this theorem for $t = 2$. Here, we give the proof for general type.

Theorem 2.2.9. *Let $\mathcal{C}$ be an $(N; n, m, d)$-code. $\mathcal{C}$ is an $\mathsf{SHF}(N; n, m, \{w_1, w_2, \ldots, w_t\})$ provided that*

$$(N - d) \sum_{1 \le i < j \le t} w_i w_j < N.$$

Proof. Let $\mathcal{C}$ be an $(N; n, m, d)$-code and $\mathcal{A} = (a_{i,j})$ an $N \times n$ array having the codewords of $\mathcal{C}$ as its columns. We interpret the array $\mathcal{A}$ as a family of N hash functions from an n-set to an m-set, i.e. each row of $\mathcal{A}$ presents a hash function. Now, consider t pairwise disjoint sets of columns of $\mathcal{A}$ denoted by $\mathsf{C}_1, \ldots, \mathsf{C}_t$, where $|\mathsf{C}_i| = w_i$ for $i = 1, \ldots, t$. As the minimum distance of code $\mathcal{C}$ is d, each two columns agree at most in $N - d$ positions. Further, the number of pairs of columns (c, c') in which $c \in \mathsf{C}_i$ and $c' \in \mathsf{C}_j$ with $i \neq j$ is $\sum_{1 \le i < j \le t} w_i w_j$. Therefore, the maximum number of rows of $\mathcal{A}$ which do not separate $\mathsf{C}_1, \ldots, \mathsf{C}_t$ is $(N - d) \sum_{1 \le i < j \le t} w_i w_j$. It means that if $N > (N - d) \sum_{1 \le i < j \le t} w_i w_j$, then there is at least one row which separates $\mathsf{C}_1, \ldots, \mathsf{C}_t$. □

A very useful property of SHFs is proved in Lemma 2.2.11. The lemma shows that by "grouping rows" of an existing SHF, we can construct a new one with less number of rows on a set having more symbols. This property is used in our proofs in the following way. In order to prove an upper bound on n in an $\mathsf{SHF}(N; n, m, \{w_1, w_2, \ldots, w_t\})$, we usually consider a certain fixed number c of rows and prove the bound for $N = c$. Then, we generalize the bound to arbitrary N using this lemma.

We first give an example to show how this grouping is done and then prove the lemma.

Example 2.2.10. The following matrix is an $\mathsf{SHF}(7; 7, 4, \{2, 2\})$.

1	2	3	4	1	2	3
1	2	3	4	2	1	4
1	2	3	4	3	4	1
1	2	3	4	4	3	2
2	3	2	3	1	1	4
2	4	1	2	3	4	3
1	1	2	2	3	4	3

We want to divide the rows into three parts, each part having three rows. We need to add two arbitrary rows:

1	2	3	4	1	2	3
1	2	3	4	2	1	4
1	2	3	4	3	4	1
1	2	3	4	4	3	2
2	3	2	3	1	1	4
2	4	1	2	3	4	3
1	1	2	2	3	4	3
1	1	1	1	1	1	1
2	2	2	2	2	2	2

We get an $\mathsf{SHF}(3; 7, 4^3, \{2, 2\})$ as follows:

$(1,1,1)$	$(2,2,2)$	$(3,3,3)$	$(4,4,4)$	$(1,2,3)$	$(2,1,4)$	$(3,4,1)$
$(1,2,2)$	$(2,3,4)$	$(3,2,1)$	$(4,3,2)$	$(4,1,3)$	$(3,1,4)$	$(2,4,3)$
$(1,1,2)$	$(1,1,2)$	$(2,1,2)$	$(2,1,2)$	$(3,1,2)$	$(4,1,2)$	$(3,1,2)$

Here we prove that grouping the rows as shown in the above example results in a separating hash family.

Lemma 2.2.11. *If there exists an* $\mathsf{SHF}(N; n, m, \{w_1, \ldots, w_t\})$*, then there exists an* $\mathsf{SHF}(\lceil \frac{N}{c} \rceil; n, m^c, \{w_1, \ldots, w_t\})$ *where* $c \geq 2$ *is an integer.*

Proof. Suppose that $\mathcal{F}$ is an $\mathsf{SHF}(N; n, m, \{w_1, \ldots, w_t\})$. Let $\mathcal{A} = (a_{i,j})$, $1 \leq i \leq N$, $1 \leq j \leq n$, be the matrix representation of $\mathcal{F}$ and $d := \lceil \frac{N}{c} \rceil$. Divide $\mathcal{A}$ into d submatrices $\mathsf{A}_1, \ldots, \mathsf{A}_d$ of size $c \times n$ (we can add arbitrary rows to $\mathcal{A}$ in order to get d submatrices with c rows).

Form a $d \times n$ array $\mathcal{B} = (b_{ij})$, with entries from Y^c as follows:

$$b_{ij} = (a_{i_1,j}, a_{i_2,j}, \ldots, a_{i_c,j}) \in Y^c, \qquad i_\ell = (i-1)c + \ell \quad \text{for } 1 \leq \ell \leq c,$$

where $(a_{i_1,j}, a_{i_2,j}, \ldots, a_{i_c,j})^T$ is the j^{th} column of A_i.

Let $\mathsf{C}_1, \ldots, \mathsf{C}_t$ be disjoint sets of columns such that $|\mathsf{C}_i| = w_i$, $1 \leq i \leq t$. Since $\mathcal{A}$ is a separating hash family of type $\{w_1, w_2, \ldots, w_t\}$, there exists some row p in $\mathcal{A}$ such that for any $1 \leq m \neq n \leq t$ we have:

$$\{a_{p,j} : j \in \mathsf{C}_m\} \cap \{a_{p,j} : j \in \mathsf{C}_n\} = \emptyset. \tag{2.1}$$

There exist integers q and $1 \leq r \leq c$ such that $p = qc + r$. Then

$$b_{q+1,j} = (a_{i_1,j}, a_{i_2,j}, \ldots, a_{i_c,j}) \qquad i_\ell = qc + \ell,\ 1 \leq \ell \leq c.$$

So, when $\ell = r$ we have $a_{i_\ell,j} = a_{i_r,j} = a_{p,j}$. It follows that in row $q+1$ of $\mathcal{B}$ the r-th component of $b_{q+1,j}$ for $1 \leq j \leq n$ is $a_{p,j}$. From (2.1) we obtain

$$\{b_{q+1,j} : j \in C_m\} \cap \{b_{q+1,j} : j \in C_n\} = \emptyset.$$

Hence, $\mathcal{B}$ is a separating hash family of type $\{w_1, w_2, \ldots, w_t\}$. □

The following Theorem on the relationship between SHFs of type $\{1, 1, 2\}$ and PHFs of strength $t = 4$ is proved in [45]. Theorem 2.2.14 gives a generalization of this result.

Theorem 2.2.12. *Let $N \leq 5$ be an integer. $\mathcal{F}$ is an $\mathsf{SHF}(N; n, m, \{1, 1, 2\})$ if and only if $\mathcal{F}$ is an $\mathsf{SHF}(N; n, m, \{1, 1, 1, 1\})$.*

The following example is the smallest SHF of type $\{1, 1, 2\}$ which is not of type $\{1, 1, 1, 1\}$.

Example 2.2.13. [45] An $\mathsf{SHF}(6; 4, 3, \{1, 1, 2\})$ which is not of type $\{1, 1, 1, 1\})$:

1	1	2	3
1	2	1	3
1	2	3	1
2	1	1	3
2	1	3	1
2	3	1	1

Theorem 2.2.12 can be generalized as follows [45].

Theorem 2.2.14. *An $\mathsf{SHF}(N; n, m, \{\overbrace{1, \ldots, 1}^{w-2}, 2\})$ is equivalent to an $\mathsf{SHF}(N; n, m, \{\overbrace{1, \ldots, 1}^{w}\})$ when $N \leq \binom{w}{2} - 1$.*

2.3 Some examples of SHFs

The aim of this section is to present constructions of SHFs and show some of the parameters for which separating hash families exist. These SHFs can be used in the recursive constructions presented in Chapter 8 to construct further separating hash

families. Moreover, they are used in the construction of cover free families which is discussed in [32].

2.3.1 $\mathsf{SHF}(4; n, m, \{1, 2\})$

Li et al. [32] proved that there does not exist an $\mathsf{SHF}(4; m^3, m, \{1, 2\})$ for a positive integer $m \geq 2$. The question here is what is the largest integer a for which an $\mathsf{SHF}(4; m^2 + a, m, \{1, 2\})$ exists. The following two examples show that for $m = 2, 3$ there exists an $\mathsf{SHF}(4; m^2 + 1, m, \{1, 2\})$.

Example 2.3.1. The following array is an optimal $\mathsf{SHF}(4; 5, 2, \{1, 2\})$ [32].

1	2	1	2	1
1	2	2	1	2
1	1	1	2	2
1	1	2	2	1

Example 2.3.2. Here is an example of an $\mathsf{SHF}(4; 10, 3, \{1, 2\})$ [32].

1	1	2	2	3	3	1	2	3	1
2	3	1	3	1	2	1	2	3	1
3	2	3	1	2	1	1	2	3	3
1	1	3	2	1	3	1	2	1	2

2.3.2 $\mathsf{SHF}(3; n, m, \{1, 1, 2\})$

In Construction 2.3.3 we present a construction for an $\mathsf{SHF}(3; n, m, \{1, 1, 1, 1\})$ with certain values of n and m. This construction presented in [45], is isomorphic to the construction given in [31] for 2-IPP codes.

Construction 2.3.3. For any positive integer t, the following construction is an $\mathsf{SHF}(3; 3t^2, t^2+2t, \{1, 1, 1, 1\})$ which according to Theorem 2.2.12 is also an $\mathsf{SHF}(3; 3t^2, t^2 + 2t, \{1, 1, 2\})$.

We divide the symbol set Y into three disjoint sets Y_1, Y_2 and Y_3 as follows:

$$\begin{aligned} Y_1 &= \{1, 2, \ldots, t\} \\ Y_2 &= \{t + 1, t + 2, \ldots, 2t\} \\ Y_3 &= \{2t + 1, \ldots, t^2 + 2t\} \end{aligned}$$

The matrix representation of the SHF consists of three different parts, each part having t^2 columns. The elements in Y_3 construct the first row of the first part, second row of the second part and third row of part three. In each part, in the other two rows we put the t^2 pairs belonging to $Y_1 \times Y_2$. These three parts are shown below:

PART I:

$2t+1$	$2t+2$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	t^2+2t
1	1	$\dots$	1	2	2	$\dots$	2	$\dots$	$\dots$	t	t	$\dots$	t
$t+1$	$t+2$	$\dots$	$2t$	$t+1$	$t+2$	$\dots$	$2t$	$\dots$	$\dots$	$t+1$	$t+2$	$\dots$	$2t$

PART II:

$t+1$	$t+2$	$\dots$	$2t$	$t+1$	$t+2$	$\dots$	$2t$	$\dots$	$\dots$	$t+1$	$t+2$	$\dots$	$2t$
$2t+1$	$2t+2$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	t^2+2t
1	1	$\dots$	1	2	2	$\dots$	2	$\dots$	$\dots$	t	t	$\dots$	t

PART III:

1	1	$\dots$	1	2	2	$\dots$	2	$\dots$	$\dots$	t	t	$\dots$	t
$t+1$	$t+2$	$\dots$	$2t$	$t+1$	$t+2$	$\dots$	$2t$	$\dots$	$\dots$	$t+1$	$t+2$	$\dots$	$2t$
$2t+1$	$2t+2$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	t^2+2t

In addition, as the above array is an $\mathsf{SHF}(3; n, m, \{1,1,1\})$ (Theorem 2.2.6) and an $\mathsf{SHF}(3; n, m, \{2,2\})$ (Theorem 2.2.7), we conclude that the array presented in Construction 2.3.3 is a 2-IPP code.

2.3.3 $\mathsf{SHF}(w+1; m^2, m, \{1, w\})$

In this section we give a construction of an $\mathsf{SHF}(w+1; m^2, m, \{1, w\})$ with $w+1 \leq m$ using orthogonal arrays.

One of the methods for constructing good classes of separating hash families is using orthogonal arrays. An *orthogonal array* $\mathsf{OA}(t, N, m)$ is an $N \times m^t$ array A with entries from a set of $m \geq 2$ symbols such that within any t rows of A every possible $t-$tuple of symbols occurs exactly once. This property is equivalent to the fact that every two columns of A agree in at most $t-1$ rows.

A classical construction of orthogonal arrays is as follows [20]. Let q be a prime power and $t \geq 2$. Let $\mathcal{P} = \{P_1, P_2, \dots, P_{q^t}\}$ be the set of all polynomials of degree at most $t-1$ over the finite field $\mathbb{F}_q$. Now let $\mathcal{R}$ be a subset of elements of $\mathbb{F}_q \cup \{\infty\}$. Define an $|\mathcal{R}| \times q^t$ array A in which the entry $\mathsf{A}(u, j)$ is $P_j(u)$ if $u \in \mathcal{R} \setminus \{\infty\}$ and is a_{t-1} when $P_j(x) = \sum_{i=0}^{t-1} a_i x^i$ and $u = \infty$. Then A is an $\mathsf{OA}(t, |\mathcal{R}|, q)$.

In the following, we present a construction for an optimal $\mathsf{SHF}(w+1; n, m, \{1, w\})$ with $n = m^2$ for $w+1 \leq m$.

Construction 2.3.4. [13] Let m be a prime power such that $w+1 \leq m$. Let $\mathsf{A} \subseteq \mathbb{F}_m$ with $|\mathsf{A}| = w+1$. Consider the classical orthogonal array $\mathsf{OA}(2, |\mathsf{A}|, m)$ which is a $(w+1) \times m^2$ array $\mathcal{A}$. Now any two different columns of $\mathcal{A}$ agree in at most one row. It follows that for given two disjoint subsets of columns C_1 and C_2 of $\mathcal{A}$ with

$|\mathsf{C}_1| = 1$ and $|\mathsf{C}_2| = w$, there is at least one row that separates C_1 and C_2. Hence $\mathcal{A}$ is an $\mathsf{SHF}(w+1; m^2, m, \{1, w\})$.

In Chapter 7, we prove Theorem 7.2.2 which provides a new upper bound on n in an SHF of type $\{1, w\}$. This theorem shows that Construction 2.3.4 gives an optimal SHF. Hence, the following theorem is obtained.

Theorem 2.3.5. *For any prime power m and any integer w with $w + 1 \leq m$, there exists an optimal* $\mathsf{SHF}(w+1; m^2, m, \{1, w\})$.

Chapter 3

Known bounds on SHFs

This chapter is a survey on the previously known upper bounds on the number of columns of separating hash families. Assume that the number of rows N, the size of alphabet set m and the type $\{w_1, w_2, \ldots, w_t\}$ are given. The question is that if there exists an SHF with these parameters, how many columns can this SHF have? The upper bound on n gives a necessary condition for the existence of an SHF with parameters $(N; n, m, \{w_1, w_2, \ldots, w_t\})$. We present existing bounds for separating hash families of different types in the following sections.

3.1 $\mathsf{SHF}(N; n, m, \{1, 1, \ldots, 1\})$

In this section we give a known bound on the number of columns of a $\mathsf{PHF}(N; n, m, t)$ which is equivalent to an $\mathsf{SHF}(N; n, m, \{\overbrace{1, 1, \ldots, 1}^{t}\})$. We begin with SHF of type $\{1, 1\}$.

It is obvious that an $N \times n$ array with m symbols is separating of type $\{1, 1\}$ if all columns are distinct and it is possible only when $n \leq m^N$. It means that the matrix representation of an $\mathsf{SHF}(N; n, m, \{1, 1\})$ consists of columns which are distinct N-tuples. Hence, we have the following theorem.

Theorem 3.1.1. *There exists an* $\mathsf{SHF}(N; n, m, \{1, 1\})$ *if and only if* $n \leq m^N$.

Theorem below gives a necessary condition for the existence of an $\mathsf{SHF}(N; n, m, \{\overbrace{1, 1, \ldots, 1}^{t}\})$ with $t \geq 3$.

Theorem 3.1.2. *[30] In an* $\mathsf{SHF}(N; n, m, \{\overbrace{1, 1, \ldots, 1}^{t}\})$ *where* $t \geq 3$, *we have* $n \leq (t-1)(m^{\lceil \frac{N}{t-1} \rceil} - 1)$.

3.2 $\mathsf{SHF}(N; n, m, \{1, w\})$

Now we consider SHFs of parameters $(N; n, m, \{1, w\})$ (w-frameproof codes) and present known bounds for this case.

Theorem 3.2.1. *[42] Assume that $w \geq 2$. If there exists an $\mathsf{SHF}(N; n, m, \{1, w\})$, then $n \leq w(m^{\lceil \frac{N}{w} \rceil} - 1)$.*

The above bound can be improved and achievable when $N \leq w$ as presented in the following:

Theorem 3.2.2. *[13] If $N \leq w$, then for each $\mathsf{SHF}(N; n, m, \{1, w\})$ we have $n \leq N(m-1)$.*

Here is a construction of an optimal $\mathsf{SHF}(N; N(m-1), m, \{1, w\})$ when $N \leq w$:

1	2	...	$m-1$	m	m	...	m	.....	m	m	...	m
m	m	...	m	1	2	...	$m-1$	.....	m	m	...	m
		$\vdots$				$\vdots$		$\ddots$			$\vdots$	
m	m	...	m	m	m	...	m	.....	1	2	...	$m-1$

3.3 $\mathsf{SHF}(N; n, m, \{w, w\})$

The following necessary condition for the existence of an $\mathsf{SHF}(N; n, m, \{w, w\})$ (w-secure frameproof codes) is proved in [42].

Theorem 3.3.1. *Suppose there exists an $\mathsf{SHF}(N; n, m, \{w, w\})$. Then*

$$n \leq m^{\lceil \frac{N}{w} \rceil} + 2w - 2.$$

This bound was improved in [45] for the type $\{2, 2\}$.

Theorem 3.3.2. *[45] If an $\mathsf{SHF}(N; n, m, \{2, 2\})$ exists, then*

$$n \leq 4m^{\lceil \frac{N}{3} \rceil} - 3.$$

In [47], Stinson et al. generalized the method used to prove Theorem 3.3.2 to get a bound on n for the type $\{w, w\}$ when $N = 2w - 1$. Here is their obtained results:

Theorem 3.3.3. *[47] If an $\mathsf{SHF}(2w-1; n, m, \{w, w\})$ exists, it holds that*

$$n \leq m + (w-1)(2w-1)(m-1).$$

By applying Lemma 2.2.11, we get the following bound on an $\mathsf{SHF}(N; n, m, \{w, w\})$.

Theorem 3.3.4. *[47] In an $\mathsf{SHF}(N; n, m, \{w, w\})$, we have*

$$n \leq (2w^2 - 3w + 2)m^{\lceil \frac{N}{2w-1} \rceil} - 2w^2 + 3w - 1.$$

3.4 $\mathsf{SHF}(N; n, m, \{w, w-1\})$

The technique used to prove the bound of Theorem 3.3.3 in [47] can also be applied for the type $\{w, w-1\}$ when $N = 2w-2$. Using this method, Stinson et al. proved also the following bound:

Theorem 3.4.1. *[47] If an* $\mathsf{SHF}(2w-2; n, m, \{w, w-1\})$ *exists, then*

$$n \leq m + (w-1)(2w-3)(m-1).$$

which results in the following theorem:

Theorem 3.4.2. *[47] If an* $\mathsf{SHF}(N; n, m, \{w, w-1\})$ *exists, then*

$$n \leq (2w^2 - 5w + 4)m^{\lceil \frac{N}{2w-2} \rceil} - 2w^2 + 5w - 3.$$

3.5 $\mathsf{SHF}(N; n, m, \{w_1, w_2\})$

In order to prove Theorems 3.3.3 and 3.4.1, Stinson et al. showed that for a large enough value of n, the matrix representation of the $\mathsf{SHF}(N; n, m, \{w, w\})$ contains a submatrix which is isomorphic to a *staircase matrix*. This concept is defined as follows:

Definition 3.5.1. [47] An (N, t)-*staircase* is a matrix S with N rows of the following form:

x_1	x_1	$*$	$*$	$*$	$*$	$*$	$*$	$*$
$*$	x_2	x_2	$*$	$*$	$*$	$*$	$*$	$*$
$\vdots$	$\vdots$	$\ddots$	$\ddots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$*$	$*$	$*$	x_{t-1}	x_{t-1}	$*$	$*$	$*$	$*$
$*$	$*$	$*$	x_t	x_t	$*$	$*$	$*$	$*$
$*$	$*$	$*$	$*$	x_{t+1}	x_{t+1}	$*$	$*$	$*$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\ddots$	$\vdots$	$\vdots$
$*$	$*$	$*$	$*$	$*$	$*$	x_{N-1}	x_{N-1}	$*$
$*$	$*$	$*$	$*$	$*$	$*$	$*$	x_N	x_N

As observed, the pattern of a staircase matrix gives a contradiction to the separating property of type $\{w, w\}$ if $N = 2w-1$ (and $\{w, w-1\}$ when $N = 2w-2$), since the even and odd indexed columns cannot be separated. Stinson and Zaverucha [48] used this forbidden configuration to prove the following bound for SHF of type $\{w_1, w_2\}$.

Theorem 3.5.2. *[48] If an* $\mathsf{SHF}(w_1 + w_2 - 1; n, m, \{w_1, w_2\})$ *with* $w_1 \geq w_2$ *exists, then*

$$n \leq m + (2w_2w_1 - w_1 - 1)(m-1).$$

This bound can be extended to N rows using Lemma 2.2.11.

Theorem 3.5.3. *[48] If an* $\mathsf{SHF}(N; n, m, \{w_1, w_2\})$ *with* $w_1 \geq w_2$ *exists, then*

$$n \leq (2w_1w_2 - w_1)m^{\lceil \frac{N}{w_1+w_2-1} \rceil} - 2w_1w_2 + w_1 + 1.$$

3.6 $\mathsf{SHF}(N; n, m, \{w_1, w_2, \ldots, w_t\})$

Let $\mathcal{F}$ be an $\mathsf{SHF}(N; n, m, \{w_1, w_2, \ldots, w_t\})$. According to Theorem 2.2.7, $\mathcal{F}$ is also an $\mathsf{SHF}(N; n, m, \{w_1, w_2 + \cdots + w_t\})$. From Theorem 3.5.3 the following bound for type $\{w_1, w_2, \ldots, w_t\}$ is obtained.

Theorem 3.6.1. *[48] Suppose $\mathcal{F}$ is an $\mathsf{SHF}(N; n, m, \{w_1, w_2, \ldots, w_t\})$ where $w_1 \leq w_2 \leq \cdots \leq w_t$. Let $u = \sum_{i=1}^{t} w_i$. Then*

$$n \leq (2w_1 - 1)(u - w_1)m^{\lceil \frac{N}{u-1} \rceil} - w_1(2u - 2w_1 + 1) + 1.$$

Independent of the result of Theorem 3.6.1, Blackburn [15] discovered the following bound for general type $\{w_1, w_2, \ldots, w_t\}$.

Theorem 3.6.2. *[15] Suppose an $\mathsf{SHF}(N; n, m, \{w_1, w_2, \ldots, w_t\})$ exists. Define $u = \sum_{i=1}^{t} w_i$. Then*

$$n \leq \binom{u}{2} m^{\lceil \frac{N}{u-1} \rceil}.$$

The exponent of m in the bounds in Theorems 3.6.1 and 3.6.2 are the same, which is claimed and discussed in [15] to be the best possible. However, Stinson and Zaverucha prove in [48] that Theorem 3.6.1 gives a stronger bound for all choices of w_i, since the coefficient of m is smaller.

The strongest known upper bound on the number of columns of an SHF with parameters $(N; n, m, \{w_1, \ldots, w_t\})$ is proved by Blackburn et al. in [17]. They show that if there exists an $\mathsf{SHF}(N; n, m, \{w_1, \ldots, w_t\})$, then n is bounded by a value which has the following general form:

$$n \leq \gamma m^{\lceil \frac{N}{u-1} \rceil} \tag{3.1}$$

where $u = \sum_{i=1}^{t} w_i$ and γ is a constant depending only on w_1, w_2, ..., w_t. It is observed that the bounds presented in Theorems 3.6.1 and 3.6.2 have the form of (3.1). Hence three different values are known for γ:

$$\begin{aligned} \gamma &= \tbinom{u}{2}, \\ \gamma &= 2(u - w_1)w_1 - w_1, \\ \gamma &= w_1 w_2 + u - w_1 - w_2. \end{aligned}$$

among which the last value is the smallest and gives the following best known bound on n in an $\mathsf{SHF}(N; n, m, \{w_1, \ldots, w_t\})$ with $u = \sum_{i=1}^{t} w_i$ where w_1 and w_2 are the smallest integers among w_i's.

Theorem 3.6.3. *[17] Suppose an $\mathsf{SHF}(N; n, m, \{w_1, \ldots, w_t\})$ exists. Define $u = \sum_{i=1}^{t} w_i$. Then*

$$n \leq (w_1 w_2 + u - w_1 - w_2)m^{\lceil \frac{N}{u-1} \rceil}$$

where $w_1, w_2 \leq w_i$ for $i = 3, \ldots, t$.

As Theorem 3.6.3 gives the best known upper bound on the number of columns of a general separating hash family, we give the proof of this theorem. In the following, first we present a lemma from [17] which gives an upper bound on an $\mathsf{SHF}(u-1; n, m, \{w_1, \ldots, w_t\})$ with $u = \sum_{i=1}^{t} w_i$. Then applying lemma 2.2.11, we obtain the bound of Theorem 3.6.3.

Lemma 3.6.4. *[17] Let $\mathcal{F}$ be an $\mathsf{SHF}(N; n, m, \{w_1, \ldots, w_t\})$ and $u = \sum_{i=1}^{t} w_i$. If $N < u$ then*

$$n \leq (w_1 w_2 + u - w_1 - w_2)(m-1) + 1.$$

Proof. Suppose that there exists an $\mathsf{SHF}(N; n, m, \{w_1, \ldots, w_t\})$, namely $\mathcal{F}$, with matrix representation $\mathcal{A} = (a_{i,j})$ where $N < u$ and

$$n = (w_1 w_2 + u - w_1 - w_2)(m-1) + 2.$$

Without loss of generality, we assume that $N = u - 1$.

In the first part of the proof, we show that there exist two columns with certain properties that we need. Then we use these two columns in the second part of the proof to derive a contradiction.

For $i = 2, \ldots, u - w_1$, let C'_i denote the set of columns having a unique element in row i. For $i = u - w_1 + 1, \ldots, u - 1$, let C'_i denote the set of columns in which the element in row i appears in row i at most w_2 times. Note that if some row has m unique elements, then $n = m$. Thus $|C'_i| \leq m - 1$ for $2 \leq i \leq u - w_1$, and $|C'_i| \leq w_2(m-1)$ when $u - w_1 + 1 \leq i \leq u - 1$.

Let $\mathcal{C}$ denote the set of columns of $\mathcal{A}$. Define $C' = \mathcal{C} \setminus (C'_2 \cup C'_3 \cup \ldots C'_{u-1})$. Observe that

$$\begin{aligned} |C'| &\geq |\mathcal{C}| - \sum_{i=2}^{u-1} |C'_i| \\ &\geq n - (u - w_1 - 1)(m-1) - (w_1 - 1)w_2(m-1) \\ &= (m-1) + 2 > m. \end{aligned}$$

So there are distinct columns $c_1, c'_1 \in C'$ that have the same element in the first row. Note, in particular, that for $i = 2, \ldots, u - w_1$, the element in column c_1 and row i appears in row i at least two times. Moreover, for $i = u - w_1 + 1, \ldots, u - 1$, the element in column c'_1 and row i appears at least $w_2 + 1$ times. These properties are used in the remaining of the proof.

Using the above columns, we construct disjoint subsets $\mathsf{C}_1, \mathsf{C}_2, \ldots, \mathsf{C}_t \subseteq \mathcal{C}$ with $|\mathsf{C}_i| \leq w_i$ that are not separated in any row. Hence $\mathcal{F}$ is not an $\mathsf{SHF}(N; n, m, \{w_1, \ldots, w_t\})$.

Let $c_1 \in \mathsf{C}_1$ and $c'_1 \in \mathsf{C}_2$. As explained above, there exists a column $c_i \neq c_1$ with $a_{i,c_i} = a_{i,c_1}$ for each $i = 2, \ldots, u - w_1$. We consider one of these columns for each row $i = 2, \ldots, u - w_1$. If for $2 \leq i < j \leq u - w_1$ we have $c_i = c_j$ then we ignore

c_j and consider only c_i. Hence the number of columns is at most $u - w_1 - 1$. Then we distribute these columns in sets $\mathsf{C}_2, \mathsf{C}_3, \ldots, \mathsf{C}_t$ such that for $i = 2, \ldots, t$ we have $|\mathsf{C}_i| \leq w_i$ and for $i \neq j$ we have $\mathsf{C}_i \cap \mathsf{C}_j = \emptyset$. It is possible, because the number of columns is at most $u - w_1 - 1$ and $\sum_{i=2}^{t} |\mathsf{C}_i| = u - w_1$.

Now consider the rows $i = u - w_1 + 1, \ldots, u - 1$. Each element a_{i,c'_1} appears at least in $w_2 + 1$ columns. So for each $i = u - w_1 + 1, \ldots, u - 1$, there exists at least one column which does not belong to C_2 and has the symbol a_{i,c'_1} in row i. If these columns are not included in C_j, $j = 3, \ldots, t$, we add them to C_1. Then $|\mathsf{C}_1| \leq w_1$. It is easily observed that the sets $\mathsf{C}_1, \mathsf{C}_2, \ldots, \mathsf{C}_t$ are not separated in any row. □

Now we apply lemma 2.2.11 on lemma 3.6.4 and prove Theorem 3.6.3 as follows:

Proof. Assume that $\mathcal{F}$ is an $\mathsf{SHF}(N; n, m, \{w_1, \ldots, w_t\})$. Let $c := \lceil \frac{N}{u-1} \rceil$. According to lemma 2.2.11, there exists an $\mathsf{SHF}(\lceil \frac{N}{c} \rceil; n, m^c, \{w_1, \ldots, w_t\})$, say $\mathcal{H}$. We show that $\lceil \frac{N}{c} \rceil \leq u - 1$. Then we use the bound in Lemma 3.6.4 on $\mathcal{H}$.

$$\begin{aligned} c = \lceil \tfrac{N}{u-1} \rceil \geq \tfrac{N}{u-1} \quad &\Rightarrow \quad \tfrac{N}{c} \leq \frac{N}{\frac{N}{u-1}} = u - 1 \\ &\Rightarrow \quad \lceil \tfrac{N}{c} \rceil \leq u - 1. \end{aligned}$$

So $\mathcal{H}$ satisfies in the conditions of Lemma 3.6.4 and we conclude that:

$$\begin{aligned} n &\leq (w_1 w_2 + u - w_1 - w_2)(m^{\lceil \frac{N}{u-1} \rceil} - 1) + 1 \\ &\leq (w_1 w_2 + u - w_1 - w_2) m^{\lceil \frac{N}{u-1} \rceil}. \end{aligned}$$

□

It should be noted that the bound presented in 3.6.3 is not the best known bound for all types of separating hash families. There exist values of t and w_i's for which better bounds are known. For example, for perfect hash families and w-frameproof codes, the bounds presented in Theorems 3.1.2 and 3.2.1 are stronger bounds. In section 3.7, we present better bounds for 2-IPP codes.

3.7 $\mathsf{SHF}(N; n, m, \{2, 2\} + \{1, 1, 1\})$

The bound presented in Theorem 3.6.3 does not provide the strongest upper bound for 2-IPP codes. In this section we present two improved bounds for this case.

Hollmann et al. proved the following bound on 2-IPP codes [31].

Theorem 3.7.1. *Suppose there exists an* $\mathsf{SHF}(N; n, m, \{2, 2\} + \{1, 1, 1\})$. *Then*

$$n \leq 3(m^{\lceil \frac{N}{3} \rceil} - 1).$$

Stinson et al. proved the bound of Theorem 3.7.2 on separating hash families of type $\{1,1,2\}$ which is a more general case of 2-IPP codes according to Lemmas 2.2.6 and 2.2.7.

Theorem 3.7.2. *[45] If an* $\mathsf{SHF}(N;n,m,\{1,1,2\})$ *exists, then*

$$n \leq 3m^{\lceil \frac{N}{3} \rceil} + 2 - 2\sqrt{3m^{\lceil \frac{N}{3} \rceil} + 1}.$$

In 6.3, we will generalize the bound in Theorem 3.7.2 to an SHF of type $\{w_1,\ldots,w_t\}$ with $t \geq 3$ (consequently $m \geq 3$) and $u \geq 4$. To prove this generalization, first we assume the case $t = 3$ and prove the following:

In an $\mathsf{SHF}(u-1;n,m,\{w_1,w_2,w_3\})$ with $u = w_1+w_2+w_3$, $w_3 \geq 2$ and $n-m \geq u-1$, we have

$$n \leq (u-1)m + 2 - 2\sqrt{3m+2}.$$

The above result is proved using induction on $u = \sum_{i=1}^{t} w_i$. According to the assumptions, the beginning step of the induction includes the case $u = 4$ which corresponds to the type $\{1,1,2\}$. i.e. the following theorem which is a special case of Theorem 3.7.2 is the base of our induction. Theorem 3.7.3 is proved in [45]. As we will use this result in the proof of Theorem 6.3.4, in the following we present the technique used by Stinson et al. for its proof.

Theorem 3.7.3. *[45] If an* $\mathsf{SHF}(3;n,m,\{1,1,2\})$ *exists, then*

$$n \leq 3m + 2 - 2\sqrt{3m+1}.$$

To prove Theorem 3.7.3, we need some preliminaries. Let $\mathcal{A}$ be the matrix representation of an arbitrary hash family and $\mathcal{C}$ denote the set of columns of $\mathcal{A}$. Hollmann et al. [31] define a multigraph $\mathcal{G}(\mathcal{A})$ on vertex set $\mathcal{C}$ as follows: The vertices c and c' are incident if there exists some row r of $\mathcal{A}$ with $\mathcal{A}(r,c) = \mathcal{A}(r,c')$.

The connected components of $\mathcal{G}(\mathcal{A})$ induce a partition of $\mathcal{A}$ into disjoint submatrices. These submatrices are called *connected components* of $\mathcal{A}$. Each connected component of $\mathcal{A}$ can be viewed as a subset of columns of $\mathcal{A}$. For every two connected components A_1 and A_2 of $\mathcal{A}$, the symbol sets of A_1 and A_2 are disjoint in the following sense: for every row r, it holds that

$$\{\mathcal{A}(r,c) : c \in \mathsf{A}_1\} \cap \{\mathcal{A}(r,c) : c \in \mathsf{A}_2\} = \emptyset.$$

The following lemma proved in [45] is used in the proof of Theorem 3.7.3.

Lemma 3.7.4. *[45] Suppose that* $\mathcal{A}$ *is the matrix representation of an* $\mathsf{SHF}(3;n,m,\{1,1,2\})$ *where* $n > m$. *Then every connected component of* $\mathcal{A}$ *is isomorphic to a matrix of type* ***I*** *or type* ***II*** *defined below.*

A matrix of type ***I*** *has the following form:*

a	a	a	$\ldots$	a
b_1	b_2	b_3	$\ldots$	b_n
c_1	c_2	c_3	$\ldots$	c_n

where the b_i's are all distinct and the c_i's are all distinct.

A matrix of type ***II*** *has the following form:*

a_1	a_2	a_3	$\ldots$	a_n
b_1	b_2	b_3	$\ldots$	b_n
c_1	c_2	c_3	$\ldots$	c_n

where the a_i's are all distinct and the n ordered pairs (b_i, c_i) are all distinct.

Now we present the proof of Theorem 3.7.3.

Proof of Theorem 3.7.3. [45] Let $\mathcal{A}$ be the matrix representation of an $\mathsf{SHF}(3; n, m, \{1, 1, 2\})$. If $n \leq m$, then $n \leq 3m + 2 - 2\sqrt{3m+1}$. Hence, we assume that $n > m$. Assume that $\mathsf{A}_1, \mathsf{A}_2, \ldots$ are the connected components of $\mathcal{A}$ which are all of type **I** or type **II** according to Lemma 3.7.4. Let n_i denote the number of columns of A_i for all i and $d_{i,r}$ denote the number of distinct elements of the set $\{\mathcal{A}(r, c) : c \in \mathsf{A}_i\}$ for $r = 1, 2, 3$.

It is easily observed that if A_i is of type **I**, then

$$d_{i,1} + d_{i,2} + d_{i,3} = 2n_i + 1$$

and if A_i is of type **II**, then

$$d_{i,1} + d_{i,2} + d_{i,3} \geq n_i + 2\sqrt{n_i}.$$

Since $n_i > 0$ is an integer, it is followed that

$$n_i + 2\sqrt{n_i} \leq 2n_i + 1,$$

and hence

$$d_{i,1} + d_{i,2} + d_{i,3} \geq n_i + 2\sqrt{n_i}$$

for all i. It is obvious that

$$m \geq \sum_i d_{i,r},$$

for $r = 1, 2, 3$ and

$$n = \sum_i n_i$$

which results in the following:

$$\begin{aligned} 3m &\geq \sum_i (d_{i,1} + d_{i,2} + d_{i,3}) \\ &\geq \sum_i (n_i + 2\sqrt{n_i}) \\ &= n + 2\sum_i \sqrt{n_i} \\ &\geq n + 2\sqrt{n}. \end{aligned}$$

This results in

$$n \leq 3m + 2 - 2\sqrt{3m+1}.$$

□

Chapter 4

SHFs with small types

In this chapter, we investigate some separating hash families with small types. We consider the values $\{3,2\}$, $\{3,3\}$ and $\{1,2\}$ for the type of separating hash families. In each case we prove some properties of SHFs which help to obtain upper bounds on n. For the type $\{1,2\}$ we also provide optimal constructions for SHFs.

In 4.1 we consider SHFs of type $\{3,2\}$ and prove an upper bound on the number of columns in this case. In 4.2 a bound on n in an $\mathsf{SHF}(N;n,m,\{3,3\})$ is proved. SHFs of type $\{1,2\}$ are studied in 4.3. The optimal constructions presented for this type show that the proved bound is achievable. In each case we present tables which show that our bounds improve all the previously known bounds for these parameters.

4.1 $\mathsf{SHF}(N;n,m,\{3,2\})$

In this section, we consider separating hash families with N rows and n columns of type $\{3,2\}$ having m symbols. The aim is to prove an upper bound on the number of columns of an SHF with these parameters. First we prove a general lemma on a $k \times n$-array which is used in our following proofs. Then we prove two properties of separating hash families of type $\{3,2\}$ with $N = 4$ rows. Using these properties, we obtain a bound on n in an $\mathsf{SHF}(4;n,m,\{3,2\})$ and generalize this bound to arbitrary N. Finally we compare our new bound with the previously presented bounds that include type $\{3,2\}$.

4.1.1 Properties

The following lemma states a simple but useful property of an array in which the number of columns is large enough. In fact, it shows that in an array with sufficiently large number of columns, there exists at least one column in which all elements are repeating in their rows of location. We use this property in the proofs of our next results.

Lemma 4.1.1. *Assume that $k, m \geq 2$ are integers. Let $\mathcal{A} = (a_{i,j})$ be a $k \times n$-array with $n > k(m-1)$. Then there exists some column ℓ in $\mathcal{A}$ in which the element $a_{i,\ell}$ is a repeating element (see Definition 2.1.9) in row i for $i = 1, \ldots, k$.*

Proof. Assume that every column of an array with k rows and n columns has at least one element which is unique in the row of its location. Let $\mathcal{C}_i$ consist of all columns having unique element in row i for $i = 1, \cdots, k$ and $\mathcal{C}$ represent the set of all columns. Then as every column has at least one unique element we conclude that $\mathcal{C} \subseteq \bigcup_{i=1}^{k} \mathcal{C}_i$. Each row can have at most $m-1$ unique elements, otherwise there are at most m columns which is less than $k(m-1)$, while $k, m \geq 2$. So

$$|\mathcal{C}| \leq \sum_{i=1}^{k} |\mathcal{C}_i| \leq k(m-1)$$

which contradicts the assumption of the lemma. □

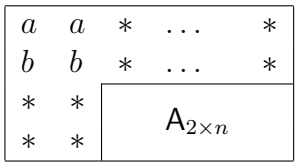
In the following, we prove two properties of an $\mathsf{SHF}(4; n, m, \{3, 2\})$.

Lemma 4.1.2 gives a necessary condition for the existence of an $\mathsf{SHF}(4; n, m, \{3, 2\})$ with $n > 2m$.

Lemma 4.1.2. *In an $\mathsf{SHF}(4; n, m, \{3, 2\})$ with $n > 2m$ every two columns have the same symbol at most in one position.*

Proof. To prove this lemma we assume that there exists an $\mathsf{SHF}(4; n, m, \{3, 2\})$ with $n > 2m$ in which two columns agree in two positions. Then we show that there exists a submatrix which is not separating of type $\{3, 2\}$.

Assume that $\mathcal{F}$ is an $\mathsf{SHF}(4; n, m, \{3, 2\})$ in which $n > 2m$ and there are two columns that agree in two positions (rows one and two). By ignoring these two columns and the first two rows we get a $2 \times n$-array A with $n > 2(m-1)$ as shown below:



According to Lemma 4.1.1, A contains a column with repeating elements which are denoted by x and y below. These columns give us one of the following configurations:

a	a	$*$	$*$	$*$
b	b	$*$	$*$	$*$
$*$	$*$	x	x	$*$
$*$	$*$	y	$*$	y

or

a	a	$*$	$*$
b	b	$*$	$*$
$*$	$*$	x	x
$*$	$*$	y	y

It is observed that in the first array the sets of columns $\mathsf{C}_1 = \{1, 3\}$ and $\mathsf{C}_2 = \{2, 4, 5\}$ are not separable and in the second array the sets $\mathsf{C}_1 = \{1, 3\}$ and $\mathsf{C}_2 = \{2, 4\}$ are

not separable. In both cases we get a contradiction to separating property of type $\{3, 2\}$. □

In the following theorem, we prove a necessary condition for the existence of an $\mathsf{SHF}(4; n, m, \{3, 2\})$ with $n > 3m - 2$.

Theorem 4.1.3. *In an* $\mathsf{SHF}(4; n, m, \{3, 2\})$ *with* $n > 3m - 2$ *the number of unique elements in each row is at most* $m - 3$.

Proof. Assume that in an $\mathsf{SHF}(4; n, m, \{3, 2\})$ with $n > 3m - 2$ there is some row with $m - 2$ unique elements. Then the remaining two elements must fill at least

$$3m - 2 + 1 - (m - 2) = 2m + 1$$

columns. So at least one element appears at least $m+1$ times. It means that there are two columns having the same symbol in two positions which is according to Lemma 4.1.2 not possible. □

4.1.2 New bound

Now, we can prove the following bound on n in an $\mathsf{SHF}(4; n, m, \{3, 2\})$ with $m > 3$.

Theorem 4.1.4. *In an* $\mathsf{SHF}(4; n, m, \{3, 2\})$ *with* $m > 3$ *we have* $n \leq 4m - 6$.

Proof. Assume that there is an $\mathsf{SHF}(4; 4m-5, m, \{3, 2\})$ with matrix representation $\mathcal{A} = (a_{i,j})$. Let $\mathcal{C}$ denote the set of columns of $\mathcal{A}$. We divide $\mathcal{C}$ into two disjoint sets $\mathcal{C}_1$ and $\mathcal{C}_2$ such that $\mathcal{C}_1$ consists of columns which have some unique element in row three or four and $\mathcal{C}_2$ consists of columns with repeating elements in rows three and four. As $m > 3$ we have $4m - 5 > 3m - 2$. Hence according to Lemma 4.1.3, $|\mathcal{C}_1| \leq 2(m - 3)$. Hence

$$|\mathcal{C}_2| = |\mathcal{C}| - |\mathcal{C}_1| \geq 4m - 5 - 2(m - 3) = 2m + 1 > 2(m - 1).$$

From Lemma 4.1.1 we conclude that there is some column in $\mathcal{C}_2$ with repeating elements in rows one and two as shown below:

a	a	$*$
b	$*$	b
$*$	$*$	c
$*$	$*$	d

(4.1)

As the third column in the above configuration belongs to $\mathcal{C}_2$, there exist two columns $c_1, c_2 \in \mathcal{C}_2$ (different from column three of (4.1)) with $a_{3,c_1} = c$ and $a_{4,c_2} = d$. According to Lemma 4.1.2, $c_1 \neq c_2$ and c_1 and c_2 are both different from the first column in the above submatrix. Hence the following two cases can happen:

(i) c_1 and c_2 are different from the second column of the array (4.1). In this case the following submatrix exists:

a	a	$*$	$*$	$*$
b	$*$	b	$*$	$*$
$*$	$*$	c	c	$*$
$*$	$*$	d	$*$	d

which is a forbidden configuration, as the column sets $\mathsf{C}_1 = \{2, 3\}$ and $\mathsf{C}_2 = \{1, 4, 5\}$ are not separable.

(ii) c_1 or c_2 is the second column of the array (4.1). Here we assume that $c_1 = 2$. The case $c_2 = 2$ is discussed in a similar way. This assumption means that a submatrix isomorphic to the following exists:

a	a	$*$	$*$
b	$*$	b	$*$
e	c	c	$*$
$*$	$*$	d	d

(4.2)

As the first column above belongs to $\mathcal{C}_2$, there exists a column c_3 different from column one of (4.2) with $a_{3,c_3} = e$. It is clear that $e \neq c$, according to Lemma 4.1.2. Hence, there are two possibilities:

1. c_3 does not belong to the set of columns in (4.2). Hence the following forbidden configuration exists:

a	a	$*$	$*$	$*$
b	$*$	b	$*$	$*$
e	c	c	$*$	e
$*$	$*$	d	d	$*$

   in which $\mathsf{C}_1 = \{1, 4\}$ and $\mathsf{C}_2 = \{2, 3, 5\}$ are not separated.

2. c_3 is the fourth column of (4.2) that results the following submatrix:

a	a	$*$	$*$
b	$*$	b	$*$
e	c	c	e
$*$	f	d	d

(4.3)

   From Lemma 4.1.2 we have $f \neq d$, as columns two and three of (4.3) have the same symbol in row three. On the other hand, column two of (4.3) belongs to $\mathcal{C}_2$ which implies that there exists a column c_4 (different from column two) such that $a_{4,c_4} = f$. If c_4 is the first column of (4.3), we get a

contradiction to Lemma 4.1.2. Hence, the following submatrix is obtained:

a	a	$*$	$*$	$*$
b	$*$	b	$*$	$*$
e	c	c	e	$*$
$*$	f	d	d	f

which is not separating of type $\{3, 2\}$, as the first and last columns are not separated from other columns. □

This bound is generalized to an SHF of type $\{3, 2\}$ with arbitrary number of rows in the following theorem.

Theorem 4.1.5. *In an* $\mathsf{SHF}(N; n, m, \{3, 2\})$ *with* $m > 3$*, we have* $n \leq 4m^{\lceil \frac{N}{4} \rceil} - 6$*.*

Proof. Assume that there exists an $\mathsf{SHF}(N; n, m, \{3, 2\})$ with $m > 3$ and $n = 4m^{\lceil \frac{N}{4} \rceil} - 5$. Let $d := \lceil \frac{N}{4} \rceil$. As explained in the proof of Lemma 2.2.11, we divide the rows of the SHF into four parts, each part containing d rows. Then we obtain an $\mathsf{SHF}(4; n, m^d, \{3, 2\})$ in which

$$n = 4m^{\lceil \frac{N}{4} \rceil} - 5 = 4m^d - 5.$$

This is a contradiction to Theorem 4.1.4. □

4.1.3 Comparison

Now we compare the bound proved in Section 4.1.2 with those previously known bounds that can be applied for type $\{3, 2\}$.

The bounds presented in Sections 3.4, 3.5 and 3.6 include type $\{3, 2\}$ as a special case. As all of these bounds (when the type is $\{3, 2\}$) together with our new bound are of the form $n \leq \gamma m^{\lceil \frac{N}{4} \rceil} - \delta$ where δ is a constant value, it is sufficient to consider $N = 4$ and show that the bound of Theorem 4.1.4 is the strongest. The result is presented in Table 4.1.

Table 4.1: Comparison of the bounds for $\mathsf{SHF}(4; n, m, \{3, 2\})$

Theorem	Obtained bound
3.4.1	$7m - 6$
3.5.2	$9m - 8$
3.6.4	$6m - 5$
4.1.4	$4m - 6$

4.2 $\mathsf{SHF}(N; n, m, \{3, 3\})$

In this section, some properties of an $\mathsf{SHF}(5; n, m, \{3, 3\})$ are proved. These properties are then used to obtain an upper bound on n in an $\mathsf{SHF}(5; n, m, \{3, 3\})$ which can be generalized to an $\mathsf{SHF}(N; n, m, \{3, 3\})$. Next the new obtained bound is compared with the previously known bounds.

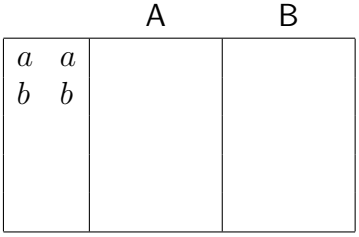
4.2.1 Properties

The following lemma gives a necessary condition for the existence of an SHF with parameters $(5; n, m, \{3, 3\})$ where $n > 4m - 2$.

Lemma 4.2.1. *In an* $\mathsf{SHF}(5; n, m, \{3, 3\})$ *with* $n > 4m - 2$ *every two columns can have the same symbol at most in one row.*

Proof. To prove this theorem we assume by contradiction that there is an $\mathsf{SHF}(5; n, m, \{3, 3\})$ with $n > 4m-2$ in which two columns have the same symbols in two rows. We then obtain two sets of columns C_1 and C_2 of cardinality three which are not separable.

By using Lemma 2.2.4 we may assume that the first two columns of the array have the same symbol in the first two rows. The separating hash family consists of three parts as shown below:



The first part includes the two columns which have the same symbols in the first two rows. Part A consists of all columns in which the elements in the last row appear at most two times in columns different from the first two and part B is the remaining of the SHF. We use the first and the last part to construct C_1 and C_2.

First we show that the number of columns of B denoted by $|\mathsf{B}|$ is at least $2(m-1)+1$. Note that if all the m elements appear at most two times in the last row, then there will be only $2m$ columns. Hence if $|\mathsf{A}|$ is the number of columns of A then $|\mathsf{A}| \leq 2(m-1)$. It implies that $|\mathsf{B}| > 4m-2-2-2(m-1) = 2m-2$. According to lemma 4.1.1, in B there exists a column with two repeating elements in rows three and four. Therefore there exists a submatrix isomorphic to the following:

$$\begin{array}{|c|c|c|c|c|}
\hline
a & a & * & * & * \\ \hline
b & b & * & * & * \\ \hline
* & * & c & c & * \\ \hline
* & * & d & * & d \\ \hline
* & * & * & e & * \\ \hline
\end{array} \tag{4.4}$$

Column four in the above configuration belongs to part B, i.e., e appears at least three times in the last row. There are two possibilities:

(i) e is the symbol in the last row of column three. In this case $\mathsf{C}_1 = \{1, 3\}$ is not separable from $\mathsf{C}_2 = \{2, 4, 5\}$ as shown below and we will not have a separating hash family of type $\{3, 2\}$.

a	a	$*$	$*$	$*$
b	b	$*$	$*$	$*$
$*$	$*$	c	c	$*$
$*$	$*$	d	$*$	d
$*$	$*$	e	e	$*$

(ii) Now assume that the symbol in the last row of column three in (4.4) is different from e. Then there exists a column not belonging to array (4.4) which has e in the last row:

a	a	$*$	$*$	$*$	$*$
b	b	$*$	$*$	$*$	$*$
$*$	$*$	c	c	$*$	$*$
$*$	$*$	d	$*$	d	$*$
$*$	$*$	$*$	e	$*$	e

As shown above, $\mathsf{C}_1 = \{1, 3, 6\}$ and $\mathsf{C}_2 = \{2, 4, 5\}$ are not separable. □

The following lemma, gives a bound on the number of unique elements in an SHF with parameters $(5; n, m, \{3, 3\})$.

Lemma 4.2.2. *In an* $\mathsf{SHF}(5; n, m, \{3, 3\})$ *with* $n > 4m - 2$*, the maximum number of unique elements in each row is* $m - 4$*.*

Proof. Assume that in some row of an $\mathsf{SHF}(5; n, m, \{3, 3\})$ with $n > 4m-2$ there are $m - 3$ unique elements. In this case each element from the three repeating elements can appear at most m times, otherwise two columns will agree in two positions which is not possible according to Lemma 4.2.1. So there are at most $m-3+3m = 4m-3$ columns which is less than $4m - 2$. It means that there can be at most $m - 4$ unique elements. □

4.2.2 New bound

Using Lemmas 4.2.1 and 4.2.2, we prove a bound on n in an $\mathsf{SHF}(5; n, m, \{3, 3\})$ with $m > 11$.

Theorem 4.2.3. *In an* $\mathsf{SHF}(5; n, m, \{3, 3\})$ *with* $m > 11$ *we have* $n < 5m - 13$*.*

Proof. Assume that there exists an $\mathsf{SHF}(5; 5m - 13, m, \{3, 3\})$ with $m > 11$ and matrix representation $\mathcal{A} = (a_{i,j})$. Let $\mathcal{C}_1$ denote the set of columns with a unique

element in row three, four or five and $\mathcal{C}_2 = \mathcal{C} \setminus \mathcal{C}_1$ where $\mathcal{C}$ is the set of columns of the SHF. We prove that there are $\mathsf{C}_1, \mathsf{C}_2 \subseteq \mathcal{C}_2$ with $|\mathsf{C}_1| = |\mathsf{C}_2| = 3$ that are not separable in the assumed SHF.

As $m > 11$ the condition of Lemma 4.2.2 holds. Hence each row can have at most $m - 4$ unique elements. Therefore $|\mathcal{C}_1| \leq 3(m-4)$ implying that

$$|\mathcal{C}_2| = |\mathcal{C}| \setminus |\mathcal{C}_1| \geq 5m - 13 - 3(m-4) = 2m - 1 > 2(m-1).$$

From Lemma 4.1.1 we conclude that there exists some column in $\mathcal{C}_2$ in which the symbols in rows one and two are repeating:

a	a	$*$
b	$*$	b
$*$	$*$	$*$
$*$	$*$	$*$
$*$	$*$	$*$

(4.5)

Columns two and three can have the same symbol at most in one position. Therefore in rows three, four and five they differ at least in two positions. By permuting the rows we can assume that they are different in rows three and four. i.e., in the following submatrix

a	a	$*$
b	$*$	b
$*$	c	d
$*$	e	f
g	$*$	$*$

(4.6)

$c \neq d$ and $e \neq f$. As the above columns belong to $\mathcal{C}_2$, there exist columns c_1, c_2, c_3 and c_4 different from column one in (4.6) (because of Lemma 4.2.1) having the same symbol as columns two and three in rows three and four. Moreover, there is a column c_5 different from columns belonging to submatrix (4.6) which has g in row five. Assume that $a_{3,c_1} = c$, $a_{3,c_2} = d$, $a_{4,c_3} = e$, $a_{4,c_4} = f$ and $a_{5,c_5} = g$. Let c_1, c_2, c_3 and c_4 be fixed. There are two possible cases:

(i) There exists a column $c_5 \neq c_1, c_3$ with $a_{5,c_5} = g$. In this case $\mathcal{A}$ has a submatrix isomorphic to the following:

a	a	$*$	$*$	$*$	$*$
b	$*$	b	$*$	$*$	$*$
$*$	c	d	c	$*$	$*$
$*$	e	f	$*$	e	$*$
g	$*$	$*$	$*$	$*$	g

in which $\mathsf{C}_1 = \{1, 4, 5\}$ and $\mathsf{C}_2 = \{2, 3, 6\}$ are not separable.

(ii) For each column c_5 with $a_{5,c_5} = g$ we have either $c_5 = c_1$ or $c_5 = c_3$. This assumption results in the following configuration with $h \neq g$:

a	a	$*$	$*$	$*$
b	$*$	b	$*$	$*$
$*$	c	d	d	$*$
$*$	e	f	$*$	f
g	$*$	$*$	$*$	h

(4.7)

h is a repeating element, as it is located in row five of a column belonging to $\mathcal{C}_2$. Let c_6 denote a column with $a_{5,c_6} = h$. c_6 is not the third column of (4.7), from Lemma 4.2.1. One of the following conditions holds:

1. c_6 is the same as the second column of (4.7). Then $\mathsf{C}_1 = \{2, 3\}$ is not separable from $\mathsf{C}_2 = \{1, 4, 5\}$ as shown below:

a	a	$*$	$*$	$*$
b	$*$	b	$*$	$*$
$*$	c	d	d	$*$
$*$	e	f	$*$	f
g	h	$*$	$*$	h

2. c_6 is the same as column four of (4.7). It implies the existence of a submatrix isomorphic to the following (the last column in the following is column c_1):

a	a	$*$	$*$	$*$	$*$
b	$*$	b	$*$	$*$	$*$
$*$	c	d	d	$*$	c
$*$	e	f	$*$	f	$*$
g	$*$	$*$	h	h	$*$

in which $\mathsf{C}_1 = \{1, 5, 6\}$ and $C_2 = \{2, 3, 4\}$ are not separable.

3. If none of the above conditions hold, then the following configuration exists in which $C_1 = \{1, 4, 5\}$ and $C_2 = \{2, 3, 6\}$ are not separable.

a	a	$*$	$*$	$*$	$*$
b	$*$	b	$*$	$*$	$*$
$*$	c	d	d	$*$	$*$
$*$	e	f	$*$	f	$*$
g	$*$	$*$	$*$	h	h

□

Here is the generalization of Theorem 4.2.3 to arbitrary number of rows.

Theorem 4.2.4. *In an* $\mathsf{SHF}(N; n, m, \{3, 3\})$ *with* $m > 11$ *we have* $n < 5m^{\lceil \frac{N}{5} \rceil} - 13$.

Proof. The proof is similar to the proof of Theorem 4.1.5.

4.2.3 Comparison

In table 4.2 we compare our new bound for type $\{3,3\}$ and $N=5$ with other known bounds for this type provided by Theorems 3.3.1, 3.3.3, 3.5.2 and 3.6.4. As explained in Section 4.1.3, this comparison is sufficient to show that our bound for an $\mathsf{SHF}(N;n,m,\{3,3\})$ with $N>11$ is the strongest bound.

Table 4.2: Comparison of the bounds for $\mathsf{SHF}(5;n,m,\{3,3\})$

Theorem	Obtained bound
3.3.1	m^2+4
3.3.3	$11m-10$
3.5.2	$15m-14$
3.6.4	$9m-8$
4.2.3	$5m-13$

4.3 $\mathsf{SHF}(N;n,m,\{1,2\})$

We study SHFs of type $\{1,2\}$ in this section. First we prove a bound on n in an $\mathsf{SHF}(N;n,m,\{1,2\})$. Then we compare this bound with the previously known bounds for this case to show that the new bound is the best known bound. Finally we present constructions of SHFs in which the number of columns achieves the new bound.

4.3.1 New bound

In this section we prove an optimal bound for $\mathsf{SHF}(N;n,m,\{1,2\})$ with arbitrary odd integer N. This bound is an improvement of the bound presented in [13] of the following form:

$$n \le m^{d+1} + O(m^d)$$

in an m-ary d-frameproof code of odd length $N=2d+1$.

Precisely, we prove the following.

Theorem 4.3.1. *For any* $\mathsf{SHF}(2d+1;n,m,\{1,2\})$ *we have* $n \le m^{d+1}$.

Proof. The following simple observation (O) is relevant for our proof. Let $\mathcal{A}$ be any $\mathsf{SHF}(2d+1;n,m,\{1,2\})$. If there are two columns of $\mathcal{A}$ agreeing in the first $(d+1)$ rows (resp. in the last $(d+1)$ rows), then the corresponding two $d-$tuples in the last d rows (resp. in the first d rows) of these two columns are unique. Since, otherwise let columns c_1 and c_2 have the same symbols in the first $d+1$ rows and column c_3

agree with c_1 in the last d rows.

a_1	a_1	$*$
a_2	a_2	$*$
$\vdots$	$\vdots$	$\vdots$
a_{d+1}	a_{d+1}	$*$
a_{d+2}	$*$	a_{d+2}
$\vdots$	$\vdots$	$\vdots$
a_{2d+1}	$*$	a_{2d+1}

It is observed that c_2 and c_3 are not separated from c_1.

Now assume there is an $\mathsf{SHF}(2d+1; m^{d+1}+1, m, \{1,2\})$. Let $\mathcal{A}$ be its matrix representation. Since $\mathcal{A}$ has $m^{d+1}+1$ columns, there are two columns agreeing in $d+1$ first rows. So, from observation (O) the d-tuples of symbols in the last d rows of these two columns are unique.

Removing these two columns from $\mathcal{A}$ gives rise to an array $\mathcal{B}$ with $m^{d+1}-1$ columns having only m^d-2 (d)-tuples of symbols distributed in the last d rows. If each d-tuple of symbols appears at most m times in the last d rows, then we can fill only $(m^d-2)m = m^{d+1}-2m < m^{d+1}-1$ columns. So, there are $m^{d+1}-1-(m^{d+1}-2m) = 2m-1$ columns in which the d-tuples of symbols in the last d rows are repeated at least $m+1$ times. This is to say that there are at least $2m-1+1 = 2m$ $(d+1)$-tuples of symbols that repeat in the last $d+1$ rows, as there are m symbols altogether. These $2m$ repeated $(d+1)$-tuples (in the last $(d+1)$ rows), provide $2m$ unique d-tuples in the first d rows by observation (O).

Removing these $2m$ columns having unique d-tuples of symbols in the first d rows from $\mathcal{A}$, gives rise to an array $\mathcal{C}$ with $m^{d+1}+1-2m$ columns having m^d-2m different d-tuples in the first d rows. If each of these m^d-2m (d)-tuples appears at most m times, then again we can fill at most $(m^d-2m)m = m^{d+1}-2m^2$ columns. So there are $m^{d+1}-2m+1-(m^{d+1}-2m^2) = 2m^2-2m+1$ columns with d-tuples in the first d rows that have to repeat at least $m+1$ times. This gives us $(2m^2-2m+1)+1 = 2m^2-2m+2$ repeated $(d+1)$-tuples in the first $d+1$ rows. Therefore, observation (O) provides $2m^2-2m+2$ unique (d)-tuples in the last d rows.

Now removing these $2m^2-2m+2$ columns from $\mathcal{B}$ we obtain an array $\mathcal{D}$ with $m^{d+1}-1-(2m^2-2m+2)$ columns having $m^d-2-(2m^2-2m+2) = m^d-2m^2+2m-4$ (d)-tuples in the last d rows. Again, if each of these d-tuples appear at most m times, only at most $(m^d-2m^2+2m-4)m = m^{d+1}-2m^3+2m^2-4m$ columns of $\mathcal{D}$ can be filled. Thus, there are $m^{d+1}-1-(2m^2-2m+2)-(m^{d+1}-2m^3+2m^2-4m) = 2m^3-4m^2+6m-3$ (d)-tuples of symbols in the last d rows repeated at least $m+1$ times. This implies that there are at least $2m^3-4m^2+6m-3+1 = 2m^3-4m^2+6m-2$ repeated $d+1$-tuples in the last $d+1$ rows. Hence, observation (O) shows that the

corresponding d-tuples in the first d rows of these $d+1$-tuples must be unique.

We see that the number of unique $d-$tuples is increasing at each step. For instance, in the next step by removing $2m^3-4m^2+6m-2$ columns containing the unique d-tuples in the first d rows from $\mathcal{C}$ we obtain an array $\mathcal{E}$ with $m^{d+1}+1-2m-(2m^3-4m^2+6m-2) = m^{d+1}-2m^3+4m^2-8m+3$ columns and $m^d-2m-(2m^3-4m^2+6m-2) = m^d+4m^2-2m^3-8m+2$ (d)-tuples for the first d rows. Again, if each of these d-tuples appear at most m times, then at most $(m^d+4m^2-2m^3-8m+2)m = m^{d+1}+4m^3-2m^4-8m^2+2m$ columns can be filled. Therefore, there are at least $m^{d+1}-2m^3+4m^2-8m+3-(m^{d+1}+4m^3-2m^4-8m^2+2m)+1 = 2m^4-6m^3+12m^2-10m+4$ repeated $d+1$-tuples in the first $d+1$ rows of $\mathcal{E}$. Hence, there are $2m^4-6m^3+12m^2-10m+4$ unique d-tuples in the last d rows. Continuing this argument after d steps will lead to a negative number of d-tuples available for a positive number of columns, a contradiction. □

4.3.2 Comparison

In Table 4.3 a comparison between the known bounds on n in an $\mathsf{SHF}(N; n, m, \{1,2\})$ with odd N and the bound proved in Theorem 4.3.1 is given.

Table 4.3: Comparison of the bounds for $\mathsf{SHF}(2d+1; n, m, \{1,2\})$

Theorem	Obtained bound
3.2.1	$2(m^{d+1}-1)$
3.4.2	$2m^{d+1}-1$
3.5.3	$2m^{d+1}-1$
3.6.3	$2m^{d+1}$
4.3.1	m^{d+1}

4.3.3 Construction of optimal $\mathsf{SHF}(2d+1; n, m, \{1,2\})$

The bound of Theorem 4.3.1 is optimal as shown in the next theorem.

Theorem 4.3.2. *If m is a prime power, then there is an optimal* SHF *with parameters $(2d+1; m^{d+1}, m, \{1,2\})$ where $2d+1 \leq m+1$.*

Proof. Let $\mathcal{A}$ be a classical $\mathsf{OA}(d+1, 2d+1, m)$. So, $\mathcal{A}$ is a $(2d+1) \times m^{d+1}$ array with entries from $\mathbb{F}_m$ and any two columns of $\mathcal{A}$ agree in at most d rows. Therefore $\mathcal{A}$ is an $\mathsf{SHF}(2d+1; m^{d+1}, m, \{1,2\})$. This separating hash family achieves the bound of Theorem 4.3.1 and is therefore optimal. □

Theorem 4.3.2 requires that m is a prime power, however if $d = 1$, we can remove this restriction.

Theorem 4.3.3. *For any integer $m \geq 2$, there is an optimal* $\mathsf{SHF}(3; m^2, m, \{1,2\})$.

Proof. It is well-known that an $\mathsf{OA}(2,3,m)$ exists for any $m \geq 2$. An easy construction of such an OA is the zero sum construction: taking all triples $[a,b,c] \in \mathbb{Z}_m^3$ with $a+b+c=0$ in $\mathbb{Z}_m$ as columns of the array. This orthogonal array is also an $\mathsf{SHF}(3;m^2,m,\{1,2\})$. □

For any integer $m \geq 2$ we have the following theorem.

Theorem 4.3.4. *Let $m = p_1^{e_1}p_2^{e_2}\ldots p_s^{e_s}$ be a prime power factorization of an integer $m \geq 2$ such that $p_1^{e_1} < p_2^{e_2} < \ldots < p_s^{e_s}$. Then there exists an optimal $\mathsf{SHF}(2d+1;m^{d+1},m,\{1,2\})$ for any positive integer d with $2d \leq p_1^{e_1}$.*

Proof. It is known by a result of Bush [24, Theorem 7.20, page 226], that there is an $\mathsf{OA}(d+1,k,m)$ for $d+1 < p_1^{e_1}$ and $k \leq p_1^{e_1}+1$. If we choose $k = 2d+1$, then an $\mathsf{OA}(d+1,2d+1,m)$ provides an optimal $\mathsf{SHF}(2d+1;m^{d+1},m,\{1,2\})$. □

4.4 Summary

In this chapter we have proved new bounds on the number of columns of SHFs with N rows, m symbols and of type $\{3,2\}$, $\{3,3\}$ and $\{1,2\}$. We summarize these new bounds in Table 4.4. In addition, for type $\{1,2\}$ some optimal constructions are provided.

Table 4.4: New bounds for SHFs of small types

SHF	New bounds
$(N;n,m,\{3,2\}), m>3$	$4m^{\lceil \frac{N}{4} \rceil}-6$
$(N;n,m,\{3,3\}), m>11$	$5m^{\lceil \frac{N}{5} \rceil}-13$
$(2d+1;n,m,\{1,2\})$	m^{d+1}

Chapter 5

SHFs of type $\{2, 2\}$

In this chapter we study SHFs of type $\{2, 2\}$ deeply. This study results in a new bound for this type and a construction for an $\mathsf{SHF}(3; n, m, \{2, 2\})$. In addition we give some examples of separating hash families which are optimal.

Our main goal is to obtain a new bound for separating hash families of type $\{2, 2\}$. As usual we first prove a new bound on n in an $\mathsf{SHF}(3; n, m, \{2, 2\})$ which is then used to derive a bound on the number of columns of an SHF with arbitrary number of rows.

In 5.1, we prove some properties of an $\mathsf{SHF}(3; n, m, \{2, 2\})$. Using these properties, we present optimal SHFs with $N = 3$ and $m = 2, 3, 4, 5, 6, 7$ in 5.2. Section 5.3 gives an upper bound on n in an $\mathsf{SHF}(N; n, m, \{2, 2\})$. We compare our new bound with the previously known bounds for the type $\{2, 2\}$ in 5.4 to show that the old bounds are improved. In 5.5 we give a construction for $\mathsf{SHF}(3; n, m, \{2, 2\})$ showing that the new bound is close to an optimal bound. Finally, in 5.6 we give some examples of separating hash families which are optimal.

5.1 Properties of an $\mathsf{SHF}(3; n, m, \{2, 2\})$

To begin, we prove some basic lemmas about the structure of an $\mathsf{SHF}(3; n, m, \{2, 2\})$. These useful lemmas are then used in the proof of next results.

The following lemma shows that in an $\mathsf{SHF}(3; n, m, \{2, 2\})$ with $n > m$ every two columns agree at most in one position.

Lemma 5.1.1. *In an* $\mathsf{SHF}(3; n, m, \{2, 2\})$ *if two columns have the same symbol in two rows, then* $n \leq m$.

Proof. Assume that $\mathcal{F}$ is an $\mathsf{SHF}(3; n, m, \{2, 2\})$ having two columns agreeing in 2 rows:

a	a
b	b
c	d

We show that each element in the last row must be unique. If c repeats in a third column, then the first column cannot be separated from the second and the third, as shown below:

a	a	$*$
b	b	$*$
c	d	c

It implies that c, and similarly d, must be unique. Now consider other elements in the last row. If some element e repeats in the last row, then we have the following submatrix:

a	a	$*$	$*$
b	b	$*$	$*$
c	d	e	e

It is easily observed that the first and the third columns are not separated from the second and the fourth. So, each element can appear in the last row at most once. It means $n \leq m$. □

According to Remark 2.2.1, we always consider separating hash families with n columns and m symbols where $n > m$. In the following, when considering an SHF with parameters $(3; n, m, \{2, 2\})$, it is always assumed that $n > m$. Consequently, every two columns agree in at most one position.

In Lemma 5.1.2, we prove that in an $\mathsf{SHF}(3; n, m, \{2, 2\})$ the following configuration is forbidden:

a	a	a	$*$	$*$
x	$*$	$*$	x	$*$
$*$	y	z	t	y(or t)

Lemma 5.1.2. *If an* $\mathsf{SHF}(3; n, m, \{2, 2\})$ *has a submatrix of the following type:*

a	a	a	$*$
x	$*$	$*$	x
$*$	y	z	t

then y, z *and* t *are unique.*

Proof. First we show that $y \neq z \neq t \neq y$. According to Lemma 5.1.1, $y \neq z$. Assume that $y = t$. Then a submatrix isomorphic to the following exists:

a	a	a	$*$
x	$*$	$*$	x
$*$	y	z	y

It is clear that the two sets of columns $\{1, 2\}$ and $\{3, 4\}$ are not separated. So $y \neq t$. (Similarly we can prove that $z \neq t$.)

Now we show that y must be unique in the last row. If y repeats, then we have the following submatrix:

a	a	a	$*$	$*$
x	$*$	$*$	x	$*$
$*$	y	z	t	y

in which the 1st and the last columns are not separated from the 2nd and the 4th. Similarly z is also unique.

Now assume that t is repeating:

a	a	a	$*$	$*$
x	$*$	$*$	x	$*$
$*$	y	z	t	t

The set of columns $\{1, 5\}$ and $\{2, 4\}$ are not separated. Hence t is also unique. □

In the next lemma, we show how we can delete some columns from an existing $\mathsf{SHF}(3; n, m, \{2, 2\})$ to obtain a separating hash family with less number of symbols.

Lemma 5.1.3. *Assume that in an* $\mathsf{SHF}(3; n, m, \{2, 2\})$ *there is some column with all repeating elements. Then there exists a submatrix such that if we delete this submatrix, we will get an* $\mathsf{SHF}(3; n_1, m_1, \{2, 2\})$ *where* m_1 *and* n_1 *satisfy in one of the following conditions:*

(a) $m_1 = m - 2$ *and* $n_1 = n - 3$,

(b) $m_1 = m - 1$ *and* $n_1 = n - 2$,

(c) or $m_1 = m - 2$ *and* $n_1 = n - 4$.

Proof. As there is a column with all repeating elements, we have a submatrix of the following form:

a	a	x	y
b	t	b	z
c	u	v	c

(5.1)

(1) First we show that either $x = y$ or x and y are unique. Assume that x repeats in a column not in the above submatrix. Then we have the following configuration:

a	a	x	$*$	x
b	$*$	b	$*$	$*$
c	$*$	$*$	c	$*$

in which the two column sets $\{1, 5\}$ and $\{3, 4\}$ are not separated. So, x can only repeat in column four of the array (5.1). It means that either $x = y$ or x and y are unique.

(2) Next we show that if $x = y$ then b and c cannot appear three times respectively in rows two and three. Suppose otherwise that b appears more than two times in the second row:

a	a	x	x	$*$
b	$*$	b	$*$	b
c	$*$	$*$	c	$*$

Then the sets of columns $\{1, 3\}$ and $\{4, 5\}$ are not separated.

The results in (1) and (2) can also be proved for the elements in rows two and three.

Now consider the following cases:

(i) None of the equalities $x = y$, $t = z$ and $u = v$ hold. So in the matrix (5.1) x, y, z, t, u and v are all unique and by deleting the columns two, three, and four we remove two symbols from each row and thus obtain an $\mathsf{SHF}(3; n-3, m-2, \{2, 2\})$.

(ii) Exactly one of the equalities in (i) holds. Assume that $x = y$ and z, t, u and v are unique. i.e., the following submatrix exists:

a	a	x	x
b	t	b	z
c	u	v	c

(5.2)

As proved in (2) above, x does not appear in other columns. Therefore by deleting columns three and four from (5.2), we remove one symbol from each row and obtain an $\mathsf{SHF}(3; n-2, m-1, \{2, 2\})$.

(iii) At least two of the equalities $x = y$, $t = z$ and $u = v$ hold. It implies the existence of an array isomorphic to the following.

a	a	x	x
b	t	b	t
c	u	v	c

(5.3)

According to (2), a, b and c do not appear in any other column in their rows of location. Hence if we delete the submatrix (5.3), two symbols are removed from each row and we obtain an $\mathsf{SHF}(3; n-4, m-2, \{2, 2\})$. □

5.2 Optimal $\mathsf{SHF}(3; n, m, \{2, 2\})$

In this section, we establish bounds for $\mathsf{SHF}(3; n, m, \{2, 2\})$ with $m = 2, 3, 4, 5, 6, 7$ and give examples to show that these bounds are optimal. We will use the bound for $m = 7$ as the induction start to prove our main theorem in the next section.

We begin with $m = 2$ and prove that if there exists an $\mathsf{SHF}(3; n, 2, \{2, 2\})$, then $n \leq 4$. Example 5.2.2 shows that this bound is achievable.

Proposition 5.2.1. *In an* $\mathsf{SHF}(3; n, 2, \{2, 2\})$ *we have* $n \leq 4$.

Proof. Assume that there exists an $\mathsf{SHF}(3; n, 2, \{2, 2\})$ with $n = 5$. So there must be some symbol that appears at least three times in row one. As there are only two symbols, at least two of these three columns – having the same symbol in the first row – agree also in the second row. It means that if some symbol appears in some row at least three times, then two columns agree in two positions. So according to Lemma 5.1.1, we can have at most $n = m = 2$ columns. We conclude that in an $\mathsf{SHF}(3; n, 2, \{2, 2\})$ each symbol can appear at most two times, hence $n \leq 4$. □

The following array is an optimal $\mathsf{SHF}(3; 4, 2, \{2, 2\})$.

Example 5.2.2. There exists an $\mathsf{SHF}(3; 4, 2, \{2, 2\})$.

a	a	b	b
a	b	a	b
a	b	b	a

Now, we prove that an $\mathsf{SHF}(3; n, 3, \{2, 2\})$ cannot have more than five columns.

Proposition 5.2.3. *In an* $\mathsf{SHF}(3; n, 3, \{2, 2\})$ *we have* $n \leq 5$.

Proof. Assume that $n = 6$. There are two possibilities:

(a) Each symbol appears in each row exactly two times. This condition implies that the symbols in all columns are repeating, hence the condition of Lemma 5.1.3 holds. As all symbols are repeating, from the proof of Lemma 5.1.3, part (1), we conclude that there exists a submatrix isomorphic to the following:

a	a	x	x
b	t	b	t
c	u	u	c

in which none of the symbols appears in the remaining of the array. By removing this submatrix we obtain an $\mathsf{SHF}(3; 2, 1, \{2, 2\})$ which is impossible.

(b) Some symbol appears in some row at least three times:

a	a	a
b	c	d
x	y	z

If b, c and d are all unique, then we have $n = 3$, because we only have three symbols. So, at least one of them repeats. If two of them repeat, then by Lemma 5.1.2, x, y and z are unique which is impossible. Hence we can assume that b is repeating (and hence y and z are unique) and c and d are unique. Hence in each row two and three there is only one symbol for four columns which is impossible by Lemma 5.1.1. □

The above proposition shows that the following array is optimal.

Example 5.2.4. The following array is an $\mathsf{SHF}(3;5,3,\{2,2\})$.

a	a	b	b	c
a	b	a	b	c
a	b	b	a	c

Here the case $m = 4$ is considered.

Proposition 5.2.5. *In an* $\mathsf{SHF}(3;n,4,\{2,2\})$*, we have* $n \leq 8$*.*

Proof. Assume that $n = 9$. In each row there must be some symbol appearing three times:

a	a	a
b	c	d
x	y	z

If b, c and d are all unique, then in the second row there is one symbol for five columns which according to Lemma 5.1.1 is not possible. Similar to Proposition 5.2.3, we assume that b is repeating and c, d, y and z are unique. So, in row two we have two symbols for seven columns. It means that one symbol should appear in row two at least in four columns. By Lemma 5.1.1, in row three we need four different symbols to fill these four columns which is not possible. □

The example below, is an $\mathsf{SHF}(3;8,4,\{2,2\})$.

Example 5.2.6. There exists an $\mathsf{SHF}(3;8,4,\{2,2\})$ which is optimal according to the previous proposition.

a	a	b	b	c	c	d	d
a	b	a	b	c	d	c	d
a	b	b	a	c	d	d	c

The following proposition gives a bound on n in an $\mathsf{SHF}(3;n,5,\{2,2\})$.

Proposition 5.2.7. *In an* $\mathsf{SHF}(3;n,5,\{2,2\})$ *we have* $n \leq 9$*.*

Proof. Assume that $n = 10$. There are two possibilities:

(a) Each symbol appears in each row exactly twice. Then by Lemma 5.1.3, we can delete four columns and remove symbols and obtain an $\mathsf{SHF}(3;6,3,\{2,2\})$ which contradicts Proposition 5.2.3.

(b) Some symbol appears in some row at least three times. i.e., a submatrix isomorphic to the following exists:

a	a	a
α	β	γ
x	y	z

Assume that α, β and γ are all unique. It means that there are two symbols to fill the remaining seven columns in row two. So one symbol appears at least four times:

a	a	a	$*$	$*$	$*$	$*$
α	β	γ	δ	δ	δ	δ
x	y	z	t	u	v	w

By Lemma 5.1.1, t, u, v and w (also x, y and z) are pairwise different. As there are only five symbols, at least two of the symbols x, y and z are from the set $\{t, u, v, w\}$. Assume that $x = t$. Then by Lemma 5.1.2, δ must be unique, which is not. It shows that at least one of the symbols α, β and γ (similarly x, y and z) is repeating. By Lemma 5.1.2, only symbols in the same column can be repeating. Assume that α and x are repeating as shown in the following matrix:

a	a	a	$*$	$*$
α	β	γ	α	δ
x	y	z	t	x

By Lemma 5.1.2, δ and t are unique. So, in row three, there are three unique symbols: y, z and t. It follows that two symbols should fill seven columns in this row. Hence one symbol, say x, has to appear four times. It gives the following submatrix:

a	a	a	$*$	$*$	$*$	$*$
α	β	γ	α	δ	$*$	$*$
x	y	z	t	x	x	x

According to Lemma 5.1.1, we need four different symbols to fill the second row of the columns containing x in row three. As the three symbols β, γ and δ are unique in row two, only two symbols remain for these positions which are not enough. A similar reason shows that the other symbol $u \neq x$ also cannot repeat four times in row three. □

The array in the following example is an optimal $\mathsf{SHF}(3; 9, 5, \{2, 2\})$.

Example 5.2.8. An $\mathsf{SHF}(3; 9, 5, \{2, 2\})$:

a	a	b	b	c	c	d	d	e
a	b	a	b	c	d	c	d	e
a	b	b	a	c	d	d	c	e

Here, the case $m = 6$ is investigated. A bound and an optimal construction are presented.

Proposition 5.2.9. *In an* $\mathsf{SHF}(3; n, 6, \{2, 2\})$ *we have* $n \leq 12$.

Proof. Assume that $n = 13$. So some symbol appears at least three times in each row:

a	a	a
α	β	γ
x	y	z

Assume that α, β and γ are unique. Then we have three symbols to fill the remaining ten columns of the second row. Hence at least one symbol has to appear at least four times:

a	a	a	$*$	$*$	$*$	$*$
α	β	γ	δ	δ	δ	δ
x	y	z	t	u	v	w

As the three symbols x, y and z and the four symbols t, u, v and w are pairwise different, one of the symbols x, y or z must be equal to one of the symbols t, u, v or w. Assume that $x = t$. Then by Lemma 5.1.2, δ must be unique which is not true. The same reasoning shows that x, y and z are not all unique. So like in Proposition 5.2.7, we assume that α and x are repeating.

a	a	a	$*$	$*$
α	β	γ	α	δ
x	y	z	t	x

By Lemma 5.1.2, δ and t are also unique. So, in row three, there are three unique symbols: y, z and t. It means that three symbols should fill ten remaining columns in this row. Then some symbol should appear four times. Two cases can happen.

(i) x appears four times. It is shown in the following submatrix:

a	a	a	$*$	$*$	$*$	$*$
α	β	γ	α	δ	λ	μ
x	y	z	t	x	x	x

By Lemma 5.1.2, λ and μ are also unique in row two. Thus there are five unique symbols in row two: β, γ, δ, λ and μ. Hence α has to appear eight times in row two, a contradiction to Lemma 5.1.1.

(ii) Now assume that some symbol, say u, $u \neq x$, appears four times in row three:

a	a	a	$*$	$*$	$*$	$*$	$*$	$*$
α	β	γ	α	δ	$*$	$*$	$*$	$*$
x	y	z	t	x	u	u	u	u

Then we need four different symbols to fill the second row of the columns having u in row three. As there are totally six symbols and three of them are unique in row two, there are not enough symbols to construct this matrix. □

The following SHF is optimal, according to the above proposition.

Example 5.2.10. An $\mathsf{SHF}(3; 12, 6, \{2, 2\})$:

a	a	b	b	c	c	d	d	e	e	f	f
a	b	a	b	c	d	c	d	e	f	e	f
a	b	b	a	c	d	d	c	e	f	f	e

□

The last case we consider is $m = 7$.

Proposition 5.2.11. *In an* $\mathsf{SHF}(3; n, 7, \{2, 2\})$, $n \leq 13$.

Proof. Assume that $n = 14$. If each symbol appears in each row exactly twice, then by Lemma 5.1.3, we can delete four columns to remove two symbols and obtain an $\mathsf{SHF}(3; 10, 5, \{2, 2\})$ which contradicts Proposition 5.2.7. So we can assume that some symbol in row one appears at least three times:

a	a	a
α	β	γ
x	y	z

We show that at most one of the symbols α, β and γ can repeat in row two.

Assume that α and β are repeating:

a	a	a	$*$	$*$
α	β	γ	α	β
x	y	z	t	u

By Lemma 5.1.2, x, y, z, t and u are all unique. So two symbols must fill nine remaining columns of the third row. It means that one symbol should appear at least five times. And we need five different symbols for the row two, above the symbol repeating five times. It is shown in the following:

a	a	a	$*$	$*$	$*$	$*$	$*$	$*$	$*$
α	β	γ	α	β	δ	η	λ	μ	ν
x	y	z	t	u	w	w	w	w	w

As there are totally seven elements, at least one of the five symbols of the set $\{\delta, \eta, \lambda, \mu, \nu\}$ belongs to the set $\{\alpha, \beta, \gamma\}$. Let $\delta = \beta$.

a	a	a	$*$	$*$	$*$	$*$	$*$	$*$	$*$
α	β	γ	α	β	β	η	λ	μ	ν
x	y	z	t	u	w	w	w	w	w

According to Lemma 5.1.2, w must be unique. But it appears five times. The contradiction is the result of the assumption that two elements of the set $\{\alpha, \beta, \gamma\}$ are repeating. Hence we conclude that at least two of the symbols from $\{\alpha, \beta, \gamma\}$, similarly $\{x, y, z\}$, are unique.

We assume that β, y, γ and z are unique. In row two, there are five symbols to fill the remaining twelve columns. Hence at least one symbol, say δ, (which can be also

α), appears at least three times in row two.

a	a	$*$	$*$	$*$
β	γ	δ	δ	δ
y	z	t	u	w

If t, u and w are all unique, then we will have five unique symbols in row three. Two symbols remain to fill nine columns implying that one of them appears at least five times. In a way similar to our above discussion, we can show that it is not possible. So at least one of the elements of the set $\{t, u, v\}$, say t, is repeating which implies that there are two unique symbols b and c in row one.

a	a	b	$*$	$*$	c
β	γ	δ	δ	δ	$*$
y	z	t	u	w	t

Now if we delete columns $\{1, 2, 3, 6\}$ we will remove two symbols from each row, getting an $\mathsf{SHF}(3; 10, 5, \{2, 2\})$, a contradiction to Proposition 5.2.7. □

Here is an example of an $\mathsf{SHF}(3; 13, 7, \{2, 2\})$.

Example 5.2.12. According to the above lemma, the following SHF is an optimal $\mathsf{SHF}(3; 13, 7, \{2, 2\})$.

a	a	b	b	c	c	d	d	e	e	f	f	g
a	b	a	b	c	d	c	d	e	f	e	f	g
a	b	b	a	c	d	d	c	e	f	f	e	g

5.3 An upper bound

In this section we present the proof of a strong bound on n in an $\mathsf{SHF}(3; n, m, \{2, 2\})$.

Theorem 5.3.1. *If an* $\mathsf{SHF}(3; n, m, \{2, 2\})$ *exists with* $m \geq 7$*, then* $n < 3(m - 2)$.

Proof. We prove the theorem using induction on m:

(i) According to Proposition 5.2.11, when $m = 7$, $n \leq 13 < 3(7 - 2)$.

(ii) Assume as an inductive hypothesis, that for $m = 7, 8, \ldots, k - 1$ the statement $n < 3(m - 2)$ is valid.

Now let $m = k$. We prove that $n < 3(m - 2)$. Assume that $n = 3m - 6$. Then there exists some symbol a appearing three times in the first row (otherwise if each symbol appears at most twice, then we have at most $2m$ columns which

is less than $3m-6$, as $m \geq 7$). By Lemma 5.1.2 we have one of the following submatrices:

$I)$

a	a	a
α	$*$	$*$
β	$*$	$*$

$II)$

a	a	a
α	β	γ
$*$	$*$	$*$

where in each case the Greek symbols are unique.

I) In this case one of the following conditions holds: (Because there are $3m-6$ columns, so in each row there exists some symbol appearing at most twice.)

I.1. There exists some symbol that does not appear in the first row. In this case by deleting the first column of the above submatrix, we obtain an $\mathsf{SHF}(3; 3m-7, m-1, \{2,2\})$, contradicting the assumption of the induction.

I.2. There exists some unique symbol in the first row. Again, if we delete the column consisting of this unique element and the first column of the above submatrix, then we get an $\mathsf{SHF}(3; 3m-8, m-1, \{2,2\})$, a contradiction to the assumption of the induction.

I.3 There exists some symbol appearing twice in the first row. By deleting the two columns containing this symbol together with the first column in the above submatrix, we get an $\mathsf{SHF}(3; 3m-9, m-1, \{2,2\})$ which is a contradiction to the assumption of the induction.

II) In this case we have a submatrix of the following form:

a	a	a
α	β	γ
x	y	z

in which α, β, γ are unique and x, y, z are repeating (otherwise if one of them is unique, then we are back to the case I)). Now we have four possibilities:

II.1. There exists some column with 3 unique symbols:
By deleting this column we remove one symbol from each row and obtain an $\mathsf{SHF}(3; 3m-7, m-1, \{2,2\})$, which is a contradiction.

II.2. There exists some column with 3 repeating symbols:
In this case we get a contradiction using Lemma 5.1.3.

II.3. There exists some column with exactly 2 unique symbols:
If these two unique symbols are in rows one and three, then by deleting this column and one further column in the submatrix II one symbol will be removed from each row, and we obtain an $\mathsf{SHF}(3; 3m-8, m-1, \{2,2\})$, which is a contradiction. So, we can assume that the unique symbols are in rows 1 and 2 (the case that the unique symbols are in rows 2 and 3 can be treated similarly) and we have the following submatrix B:

$$B = \begin{array}{|c|c|}\hline u & * \\ \hline v & * \\ \hline c & c \\ \hline \end{array}$$

where u and v are unique. If there exists a symbol which does not appear in the last row, then by deleting the first column of B, we remove one symbol from each row and get an $\mathsf{SHF}(3; 3m-7, m-1, \{2,2\})$ which is a contradiction. If there exists a unique symbol in the last row, then by deleting the column containing this unique symbol and the first column of B we get an $\mathsf{SHF}(3; 3m-8, m-1, \{2,2\})$, again a contradiction.

Otherwise all of the m symbols should appear at least twice in the last row. As there are $3m-6$ columns ($< 3m$), some element should appear exactly twice. By deleting the columns containing this symbol and the first column of B we get an $\mathsf{SHF}(3; 3m-9, m-1, \{2,2\})$, which contradicts the assumption of the induction.

II.4. Every column contains exactly one unique symbol:
If all the unique symbols are in one row, then we can have at most m columns. So at least two rows contain the unique symbols. If only two rows contain the unique symbols, then $n < 2m$. Therefore, each row contains at least one unique symbol. It means that we can delete three columns having unique symbols in three different rows and get an $\mathsf{SHF}(3; 3m-9, m-1, \{2,2\}$ which again contradicts the assumption of the induction. □

Starting with the bound $n < 3(m-2)$ for an $\mathsf{SHF}(3; n, m, \{2,2\})$ proved above we now use Lemma 2.2.11 to generalize this bound for any arbitrary number of rows N.

Theorem 5.3.2. *If an* $\mathsf{SHF}(N; n, m, \{2,2\})$ *exists with* $m^{\lceil \frac{N}{3} \rceil} \geq 7$*, then* $n < 3m^{\lceil \frac{N}{3} \rceil} - 6$.

Proof. Assume there is an $\mathsf{SHF}(N; n, m, \{2,2\})$. Let $c := \lceil \frac{N}{3} \rceil$. Then by Lemma 2.2.11 there is an $\mathsf{SHF}(3; n, m^{\lceil \frac{N}{3} \rceil}, \{2,2\})$. By Theorem 6.3.1 we obtain $n < 3(m^{\lceil \frac{N}{3} \rceil} - 2)$. □

As $2-IPP$ codes are SHF of type $\{2,2\}$ and $\{1,1,1\}$ we obtain the following bound for $2-IPP$ codes as well.

Corollary 5.3.3. *For any* $2-IPP$ *code of length* N *and of* n *codewords over an alphabet of size* m *with* $m^{\lceil \frac{N}{3} \rceil} \geq 7$ *we have*

$$n < 3m^{\lceil \frac{N}{3} \rceil} - 6.$$

The bound in Corollary 5.3.3 is an improvement of the bound presented in Theorem 3.7.1.

5.4 Comparison

Table 5.1 gives a comparison between the previously known bounds on SHFs of type $\{2,2\}$ with $N=3$ and the new bound in Theorem 5.3.1. This comparison shows that the bound of Theorem 5.3.2 is the strongest bound on n in an $\mathsf{SHF}(N;n,m,\{2,2\})$.

Table 5.1: Comparison of the bounds for $\mathsf{SHF}(3;n,m,\{2,2\})$

Theorem	Obtained bound
3.3.1	m^2+2
3.3.2	$4m-3$
3.3.3	$4m-3$
3.5.2	$6m-5$
3.6.4	$4m-3$
5.3.1	$3m-6$

5.5 Constructions of $\mathsf{SHF}(3;n,m,\{2,2\})$

In this section we provide constructions for SHFs of type $\{2,2\}$ and discuss about their optimality.

Using Example 5.2.2 we see that an $\mathsf{SHF}(3;2m,m,\{2,2\})$ for even values of m can always be constructed by taking the union of $m/2$ copies of $\mathsf{SHF}(3;4,2,\ \{2,2\})$, constructed on any $m/2$ pairwise disjoint 2-subsets of the symbol set. In fact, this type of "doubling construction" provides optimal separating hash families for small values of m. But, as soon as m is large enough, say $m \geq 32$, then the construction in this section becomes superior. In addition, when m is odd, we can construct an $\mathsf{SHF}(3;2m-1,m,\{2,2\})$ in the following way. We construct an $\mathsf{SHF}(3;2m-2,m-1,\{2,2\})$ using the construction just described on $m-1$ symbols. Then we add a column having the same symbol in three rows using the remaining symbol.

The following theorem is obtained.

Theorem 5.5.1. *For any odd integer m there is an $\mathsf{SHF}(3;2m-1,m,\{2,2\})$ and for an even m there exists an $\mathsf{SHF}(3;2m,m,\{2,2\})$.*

Now we present a general construction of a good class of $\mathsf{SHF}(3;n,m,\{2,2\})$. This construction is a generalization of Construction 2.3.3. We believe that for large values of m the number of columns n obtained from this construction is close to an optimal bound. To be more precise, the construction provides separating hash families with roughly $n \cong 3(m-2\lfloor\sqrt{m}\rfloor)$ columns. Thus $\lim_{m\to\infty} \tilde{n}/m = 3$, where $\tilde{n}$ is a value of n such that an $\mathsf{SHF}(3;n,m,\{2,2\})$ exists. This implies that $c=3$ is asymptotically

the best possible minimum value for the constant c such that $n < c(m-k)$ for any fixed number $k > 0$.

Let $m \geq 2$ be an integer. We write $m = m_1 + 2m_2$, where $m_2 = \lfloor \sqrt{m} \rfloor$. Let

$$V = V_1 \cup V_2 \cup V_3$$

be a set of m symbols consisting of a union of three disjoint sets $V_1 = \{x_1, \ldots, x_{m_1}\}$, $V_2 = \{y_1, \ldots, y_{m_2}\}$, and $V_3 = \{z_1, \ldots, z_{m_2}\}$.

Construction

Let $r \geq 1$, $\delta, u \geq 0$ be integers such that
a) $r \leq m_2$ and
b) $0 \leq m_1 - r(m_2 - \delta) := u \leq m_2$, i.e., $r(m_2 - \delta) + u = m_1$.
Define the following $(1 \times m_1)$ arrays:

$\mathsf{X} = [x_1 \ldots x_{m_1}]$

$\mathsf{Y}_1 = [\underbrace{y_1 \ldots y_1}_{r} \ \underbrace{y_2 \ldots y_2}_{r} \ \cdots \ \underbrace{y_{m_2-\delta} \ldots y_{m_2-\delta}}_{r} \ \underbrace{y_{m_2-\delta+1} \ldots y_{m_2-\delta+1}}_{u}]$

$\mathsf{Y}_2 = [\underbrace{y_1 y_2 \ldots y_r \quad y_1 y_2 \ldots y_r \quad \cdots \quad y_1 y_2 \ldots y_r}_{m_2-\delta} \quad y_1 y_2 \ldots y_u]$

$\mathsf{Z}_1 = [\underbrace{z_1 \ldots z_1}_{r} \ \underbrace{z_2 \ldots z_2}_{r} \ \cdots \ \underbrace{z_{m_2-\delta} \ldots z_{m_2-\delta}}_{r} \ \underbrace{z_{m_2-\delta+1} \ldots z_{m_2-\delta+1}}_{u}]$

$\mathsf{Z}_2 = [\underbrace{z_1 z_2 \ldots z_r \quad z_1 z_2 \ldots z_r \quad \ldots \quad z_1 z_2 \ldots z_r}_{m_2-\delta} \quad z_1 z_2 \ldots z_u]$

Now define an $3 \times 3m_1$ array $\mathcal{A}$.

$$\mathcal{A} = \begin{array}{|c|c|c|} \hline \mathsf{X} & \mathsf{Y}_1 & \mathsf{Z}_1 \\ \hline \mathsf{Y}_1 & \mathsf{X} & \mathsf{Z}_2 \\ \hline \mathsf{Y}_2 & \mathsf{Z}_2 & \mathsf{X} \\ \hline \end{array}$$

We show that $\mathcal{A}$ is an $\mathsf{SHF}(3; 3m_1, m, \{2, 2\})$.
For simplicity we call the set of first m_1 columns of $\mathcal{A}$ S_1, of next m_1 columns S_2 and of last m_1 columns S_3. The following observation is useful:
-Two different columns from each S_i, $i = 1, 2, 3$, agree in at most one row.
-Two columns from two different S_i and S_j do not agree in any row.
Now let $\mathsf{C}_1 = \{c_1, c_2\}$ and $\mathsf{C}_2 = \{d_1, d_2\}$ be two disjoint sets of columns of $\mathcal{A}$. We need to consider the following cases.

a) C_1 is a subset of an S_i, so without loss of generality, we may assume $\mathsf{C}_1 \subseteq S_1$. Then the first row of $\mathcal{A}$ will separate C_1 from any C_2.
b) C_1 and C_2 are not subsets of any S_i. So we may assume $c_1 \in S_1$ and $c_2 \in S_2$. Now if $d_1 \in S_1$ and $d_2 \in S_3$, then from the above observation d_2 separates from C_1 at any row and d_1 separates from C_1 in at least two rows. Thus C_1 and C_2 are separated.
The remaining case is: $d_1 \in S_1$ and $d_2 \in S_2$. Again the above observation states that

d_1 separates from C_1 in at least two rows and also d_2 separates from C_1 in at least two rows. It follows that there is at least one row separating C_1 and C2. Hence $\mathcal{A}$ is an $\mathsf{SHF}(3; 3m_1, m, \{2, 2\})$.

Especially, if m is of the form $m = v^2 + 2v$, we then choose $r = v$, $\delta = u = 0$ and we obtain the following result. This special case is Construction 2.3.3.

Proposition 5.5.2. *There is an* $\mathsf{SHF}(3; 3v^2, v^2 + 2v, \{2, 2\})$ *for any integer* $v \geq 1$.

Observe that the construction above still leaves room for slight improvement depending on the values of m. For instance, assume $m = v^2$. Then we have $m_2 = v$, $m_1 = v^2 - 2v$. If we choose $r = v - 1$, $\delta = 2$ and $u = v - 2$, then one symbol in V_2 and one in V_3 are not used in the construction. These two free symbols are then used to form an $\mathsf{SHF}(3; 4, 2, \{2, 2\})$ by Example 5.2.2. In this way we have constructed 4 more columns. Hence we have the following.

Proposition 5.5.3. *There is an* $\mathsf{SHF}(3; 3(v^2 - 2v) + 4, v^2, \{2, 2\})$ *for any integer* $v \geq 2$.

5.6 Constructions of optimal SHFs of type $\{2, 2\}$ and $m = 2$

For given N and m determining an optimal bound for n in an $\mathsf{SHF}(N; n, m, \{2, 2\}$ appears to be a challenging problem. In this section we present two separating hash families and prove that they are optimal. The presented SHFs are both of type $\{2, 2\}$. It is observed that the number of columns in these optimal SHFs is less than the bound proved in 5.3.2.

We mention that the symbols used in the following configurations are not necessarily distinct. As there are only two symbols, we denote two different symbols as complements, i.e. a and $\overline{a}$ present two different symbols. However, a and b can be the same.

5.6.1 $\mathsf{SHF}(5; n, 2, \{2, 2\})$

We show in the following that an $\mathsf{SHF}(5; 5, 2, \{2, 2\})$ does not exist. i.e. in an $\mathsf{SHF}(5; n, 2, \{2, 2\})$ we have $n \leq 4$. Then we present an $\mathsf{SHF}(5; 4, 2, \{2, 2\})$ which is optimal.

Lemma 5.6.1. *There exists no* $\mathsf{SHF}(5; 5, 2, \{2, 2\})$.

Proof. Assume that $\mathcal{F}$ is an $\mathsf{SHF}(5; 5, 2, \{2, 2\})$. One of the following four conditions can hold:

(1) There are two columns agreeing in four positions.

(2) There are two columns agreeing in three positions.

(3) There are two columns agreeing in two positions.

(4) There are two columns agreeing in one positions.

We investigate the above cases and show that under these conditions the separating property does not satisfy.

(1) This can be concluded from (2) in the following.

(2) Assume that two columns agree in three positions as shown below:

a	a
b	b
c	c
d	$\overline{d}$
e	$\overline{e}$

Then the two pairs (d, e) and $(\overline{d}, \overline{e})$ cannot appear in the last two rows any more. Otherwise, we will not have the separating property. So the possible choices for the last two rows are $(d, \overline{e})$ and $(\overline{d}, e)$. In order to fill the remaining three columns, one of them should appear at least two times giving the following submatrix:

a	a	$*$	$*$
b	b	$*$	$*$
c	c	$*$	$*$
d	$\overline{d}$	d	d
e	$\overline{e}$	$\overline{e}$	$\overline{e}$

It is observed that the first and the last columns are not separated from the second and the third.

(3) Let two columns agree in two positions:

a	a
b	b
c	$\overline{c}$
d	$\overline{d}$
e	$\overline{e}$

As the symbols in the last three rows of the above columns are complements, these two columns cannot be separated from other columns in the last three rows. Hence in order to have the separating property between the set containing the above two columns and other sets of columns, in the first two rows of the three remaining columns we cannot have the pair (a, b). In addition we cannot have two pairs of the form (a, x) and (y, b) in two columns simultaneously, where x and y are arbitrary elements. The pairs allowed to fill the first two rows of the three remaining columns are either $\big((\overline{a}, \overline{b})$ together with $(a, \overline{b})\big)$ or

$\big((\overline{a},\overline{b})$ together with $(\overline{a},b)\big)$. It means that we will have one of the following submatrices:

(i)

a	a	$\overline{a}$	$\overline{a}$	$\overline{a}$
b	b	$*$	$*$	$*$
c	$\overline{c}$	$*$	$*$	$*$
d	$\overline{d}$	$*$	$*$	$*$
e	$\overline{e}$	$*$	$*$	$*$

or (ii)

a	a	$*$	$*$	$*$
b	b	$\overline{b}$	$\overline{b}$	$\overline{b}$
c	$\overline{c}$	$*$	$*$	$*$
d	$\overline{d}$	$*$	$*$	$*$
e	$\overline{e}$	$*$	$*$	$*$

Assume that the first case happens (The second case can be discussed in a similar way). As the elements in the last three rows of the first two columns are complements, the third column agrees with column one or two at least in two positions from the last three rows. We assume that it agrees with column one in rows three and five:

a	a	$\overline{a}$	$\overline{a}$	$\overline{a}$
b	b	$*$	$*$	$*$
c	$\overline{c}$	c	$*$	$*$
d	$\overline{d}$	$*$	$*$	$*$
e	$\overline{e}$	e	$*$	$*$

According to item (2), two columns cannot agree in three positions. Hence, column three is as shown below:

a	a	$\overline{a}$	$\overline{a}$	$\overline{a}$
b	b	$\overline{b}$	$*$	$*$
c	$\overline{c}$	c	$*$	$*$
d	$\overline{d}$	$\overline{d}$	$*$	$*$
e	$\overline{e}$	e	$*$	$*$

In the above submatrix, columns two and three have the same symbol in one position. In item (4), we will show that it leads to a contradiction.

(4) Two columns agree in one position:

a	a
b	$\overline{b}$
c	$\overline{c}$
d	$\overline{d}$
e	$\overline{e}$

The set of the first two columns can be separated from every other set of columns only in the first row, as these two columns have complements in other rows. It means the the element in the first row of the columns 3-5 in the above

configuration must be $\overline{a}$.

a	a	$\overline{a}$	$\overline{a}$	$\overline{a}$
b	$\overline{b}$	$*$	$*$	$*$
c	$\overline{c}$	$*$	$*$	$*$
d	$\overline{d}$	$*$	$*$	$*$
e	$\overline{e}$	$*$	$*$	$*$

Now in the second row of the last three columns one element must repeat which means that two columns, for example 3 and 4, agree in row two. From item (2), they must be different in rows 3-5:

a	a	$\overline{a}$	$\overline{a}$	$\overline{a}$
b	$\overline{b}$	x	x	$*$
c	$\overline{c}$	y	$\overline{y}$	$*$
d	$\overline{d}$	z	$\overline{z}$	$*$
e	$\overline{e}$	t	$\overline{t}$	$*$

Column 5 must agree with column 3 or 4 at least in two positions from the three last rows, i.e. two columns agree in three rows which is not allowed from (2). □

The example below presents an optimal SHF.

Example 5.6.2. It is easy to observe that the following matrix represents an SHF with parameters $(5, 4, 2, \{2, 2\})$. According to Lemma 5.6.1 it is an optimal separating hash family, since there exists no $\mathsf{SHF}(5; n, 2, \{2, 2\})$ with $n > 4$.

1	1	0	0
1	0	1	0
1	0	0	1
1	0	0	0
0	1	0	1

□

Remark 5.6.3. *If we apply Theorem 5.3.2 on an* $\mathsf{SHF}(5; n, 2, \{2, 2\})$ *we obtain* $n \leq 6$. *Example 5.6.2 shows that this bound does not provide the achievable value of* n *in this case.*

5.6.2 $\mathsf{SHF}(6; n, 2, \{2, 2\})$

In this section, we examine SHFs with parameters $(6; n, 2, \{2, 2\})$ and prove that an SHF with these parameters can have at most five columns. Then we present an $\mathsf{SHF}(6; 5, 2, \{2, 2\})$.

Lemma 5.6.4. *There is no* $\mathsf{SHF}(6; 6, 2, \{2, 2\})$.

Proof. To prove this lemma, we assume that an $\mathsf{SHF}(6;6,2,\{2,2\})$ exists and get contradiction.

Assume that $\mathcal{F}$ is an $\mathsf{SHF}(6;6,2,\{2,2\})$. There are four possibilities.

(1) There are two columns having the same symbols in five or four rows. The fact that such two columns cannot exist is a result of (2).

(2) Two columns agree in three positions:

a	a
b	b
c	c
d	$\overline{d}$
e	$\overline{e}$
f	$\overline{f}$

Column three is equal to one of the columns one or two at least in two positions from rows three, four and five (here we consider column 1). Assume that column three and one have the same symbols in the last three positions.

a	a	$*$
b	b	$*$
c	c	$*$
d	$\overline{d}$	d
e	$\overline{e}$	e
f	$\overline{f}$	e

It is observed that columns two and three are not separable from column one. Hence we can assume that column three is equal to column one in rows four and five.

a	a	$*$
b	b	$*$
c	c	$*$
d	$\overline{d}$	d
e	$\overline{e}$	e
f	$\overline{f}$	$\overline{f}$

If $\overline{f}$ appears in column four, then columns two and three are not separable from one and four. So we have the following construction:

a	a	$*$	$*$	$*$	$*$
b	b	$*$	$*$	$*$	$*$
c	c	$*$	$*$	$*$	$*$
d	$\overline{d}$	d	$*$	$*$	$*$
e	$\overline{e}$	e	$*$	$*$	$*$
f	$\overline{f}$	$\overline{f}$	f	f	f

As there are only two elements, at least two of the columns four, five and six agree in row five:

a	a	$*$	$*$	$*$	$*$
b	b	$*$	$*$	$*$	$*$
c	c	$*$	$*$	$*$	$*$
d	$\overline{d}$	d	$*$	$*$	$*$
e	$\overline{e}$	e	x	x	$*$
f	$\overline{f}$	$\overline{f}$	f	f	f

Now consider row four of columns four and five. Either they agree in this row or they are different.

a	a	$*$	$*$	$*$	$*$
b	b	$*$	$*$	$*$	$*$
c	c	$*$	$*$	$*$	$*$
d	$\overline{d}$	d	y	z	$*$
e	$\overline{e}$	e	x	x	$*$
f	$\overline{f}$	$\overline{f}$	f	f	f

If $y = z = d$ or $y = z = \overline{d}$ or $y = d$ and $z = \overline{d}$ then $\{1, 5\}$ and $\{2, 4\}$ are not separable. If $y = \overline{d}$ and $z = d$ then $\{1, 4\}$ and $\{2, 5\}$ are not separable.

(3) Two columns agree in two positions:

a	a
b	b
c	$\overline{c}$
d	$\overline{d}$
e	$\overline{e}$
f	$\overline{f}$

We know that every two columns cannot agree in three positions. As there are only two elements, so columns three to six in rows three to six should agree exactly in two rows with column one and exactly in two rows with column two. It implies that these columns are all different from columns one and two in the first two rows:

a	a	$\overline{a}$	$\overline{a}$	$\overline{a}$	$\overline{a}$
b	b	$\overline{b}$	$\overline{b}$	$\overline{b}$	$\overline{b}$
c	$\overline{c}$	$*$	$*$	$*$	$*$
d	$\overline{d}$	$*$	$*$	$*$	$*$
e	$\overline{e}$	$*$	$*$	$*$	$*$
f	$\overline{f}$	$*$	$*$	$*$	$*$

Now assume that column three agrees with column one in rows three and four and agrees with column two in rows five and six. Column four should agree with column two in rows three and four and with column one in rows five and six. Without loss of generality assume that column five agrees with column one in

row three. It means that columns three and five agree in three positions which is not possible.

(4) Two columns agree in one position:

a	a
b	$\overline{b}$
c	$\overline{c}$
d	$\overline{d}$
e	$\overline{e}$
f	$\overline{f}$

Column three should agree with one of these two columns at least in three positions which is a contradiction. □

The array presented in the following example is an optimal SHF.

Example 5.6.5. Lemma 5.6.4 shows that the following array is an optimal $\mathsf{SHF}(6; 5, 2, \{2, 2\})$.

0	0	1	1	1
1	0	0	0	1
0	1	0	0	1
1	1	0	1	0
0	1	1	0	0
1	0	1	0	0

Moreover, according to Lemma 5.6.1, there exists no $\mathsf{SHF}(5; 5, 2, \{2, 2\})$. It means that in an $\mathsf{SHF}(N; 5, 2, \{2, 2\})$ we have $N \geq 6$. Hence the above array has the minimum possible value of rows and is also optimal in this sense.

Remark 5.6.6. *By applying Theorem 5.3.2 on an* $\mathsf{SHF}(6; n, 2, \{2, 2\})$ *we have* $n \leq 6$ *which is larger than the optimal value of* n.

Chapter 6

Bounds for SHFs of general type

The problem of finding upper bounds on SHFs of general type has been an interesting topic for researchers recently. In 2007, Blackburn [15] proved that the number of columns of an $\mathsf{SHF}(N; n, m, \{w_1, w_2, \ldots, w_t\})$ with $u = \sum_{i=1}^{t} w_i$ satisfies in the following inequality:

$$n \leq \binom{u}{2} m^{\lceil \frac{N}{u-1} \rceil}.$$

To prove the above bound, Blackburn used the idea of labelled graphs [17].

Stinson and Zaverucha [48], used the staircases as forbidden configurations to prove the following bound on n:

$$n \leq (2w_1 - 1)(u - w_1)m^{\lceil \frac{N}{u-1} \rceil} - w_1(2u - 2w_1 + 1) + 1.$$

Later, in 2008, Blackburn et al. [17] proved the following bound (Lemma 3.6.4):

$$n \leq (w_1 w_2 + u - w_1 - w_2)m^{\lceil \frac{N}{u-1} \rceil}$$

To prove the above two bounds, it is assumed that $w_1, w_2 \leq w_i$ for $i = 3, \ldots, t$.

In this chapter, we present three new upper bounds on the number of columns of SHFs of general type which are stronger than the above bounds. The first bound is valid for all types of general SHFs. The second bound, which improves the first one, is valid for $t \geq 3$. These two bounds are presented in [7]. The third bound satisfies for an $\mathsf{SHF}(u - 1; n, m, \{w_1, \ldots, w_t\})$ with $u \geq 4$ and $t \geq 3$. Our techniques for proving the first and the third bounds are completely different from the methods used until now in literature. In particular, we believe that Lemma 6.1.1 can be used to obtain more new results about SHFs.

6.1 First general bound

The aim of this section is to show that in an $\mathsf{SHF}(N; n, m, \{w_1, w_2, \ldots, w_t\})$ with $u = \sum_{i=1}^{t} w_i$ we have

$$n \leq (u-1)m^{\lceil \frac{N}{(u-1)} \rceil}.$$

To prove this result we first show that the bound is valid for $N = u - 1$ and $t = 2$. Then we prove it for arbitrary value of N and $t = 2$. Finally, using Theorem 2.2.7 we conclude that for every integer $t \geq 2$ the bound holds.

The idea used to obtain the bound of this section is a new idea based on induction. Lemma 6.1.1 is a useful lemma necessary to prove Lemma 6.1.2. In Lemma 6.1.1 we show how we can remove some rows and columns from an existing separating hash family and get a new separating hash family with new parameters.

Lemma 6.1.1. *Suppose there exists an* $\mathsf{SHF}(N; n, m, \{w_1, w_2\})$ *with* $n - m \geq w_1 + w_2 - 1$ *and* $w_2 \geq 2$. *Then there exists an* $\mathsf{SHF}(N-1; n_1, m, \{w_1, w_2 - 1\})$ *with* $n_1 \geq n - m$.

Proof. According to Remark 2.2.1, we have $N - 1 \geq 1$ and $n_1 \geq 1$.

Let $\mathcal{A}$ be the matrix representation of an $\mathsf{SHF}(N; n, m, \{w_1, w_2\})$ with $w_2 \geq 2$. Let $m_1 \leq m$ denote the number of symbols that appear in the first row of $\mathcal{A}$. Since permuting the columns of $\mathcal{A}$ does not change the separation property (according to Lemma 2.2.4), we may assume that the first row of $\mathcal{A}$ has pairwise different symbols in the first m_1 columns. Let $\mathcal{A}_1$ denote the $(N-1) \times (n - m_1)$ matrix obtained from $\mathcal{A}$ by ignoring the first row and the first m_1 columns of $\mathcal{A}$. Set $n_1 := n - m_1$. Then $n_1 \geq n - m$. We claim that $\mathcal{A}_1$ is an $\mathsf{SHF}(N-1; n_1, m, \{w_1, w_2 - 1\})$.

Assume that $\mathcal{A}_1$ is not an $\mathsf{SHF}(N-1; n_1, m, \{w_1, w_2-1\})$. Then there are two column sets C_1 and C_2 with $|\mathsf{C}_1| = w_1$ and $|\mathsf{C}_2| = w_2 - 1$, that are not separated in any row of $\mathcal{A}_1$. Now consider the sets of columns C_1 and C_2 in matrix $\mathcal{A}$. Let a be a symbol appearing in some column of C_1 in the first row of $\mathcal{A}$. Then in the first m_1 columns of $\mathcal{A}$ there is a column c having symbol a in the first row. Add this column c to C_2. Now it is easily checked that C_1 and $\mathsf{C}_2 \cup \{c\}$ are not separated in $\mathcal{A}$, which contradicts the separating property of $\mathcal{A}$. □

Using Lemma 6.1.1, we prove in the following that a necessary condition for the existence of an $\mathsf{SHF}(u-1; n, m, \{w_1, w_2\})$ with $u = w_1 + w_2$ is that $n \leq (u-1)m$.

Theorem 6.1.2. *Suppose there exists an* $\mathsf{SHF}(u-1; n, m, \{w_1, w_2\})$ *where* $u = w_1 + w_2$. *Then* $n \leq (u-1)m$.

Proof. We prove the theorem using induction on u. Note that $u \geq 2$, as it is the sum of two positive integers. Let $\mathcal{A}$ be the matrix representation of an $\mathsf{SHF}(u - 1; n, m, \{w_1, w_2\})$.

(i) First we show that the statement is true for $u = 2$. Then $w_1 = w_2 = 1$ and $\mathcal{A}$ is a $1 \times n$ matrix. Hence, in order to have an SHF of type $\{1, 1\}$, all n symbols in the unique row of $\mathcal{A}$ must be pairwise different, i.e. $n \leq m$.

(ii) Now assume, as an inductive hypothesis, that the statement $n \leq (u-1)m$ is valid for $u = k-1 \geq 2$. i.e. in an $\mathsf{SHF}(k-2; n, m, \{w_1', w_2'\})$ with $k-1 = w_1' + w_2'$ we have $n \leq (k-2)m$.

Suppose now that $u = k$ and there exists an $\mathsf{SHF}(k-1; n, m, \{w_1, w_2\})$ such that $n > (k-1)m$, where $k = w_1 + w_2$. As $k \geq 3$, we may assume $w_2 \geq 2$. (If $w_1 = w_2 = 1$, then $u = k = 2$.) From $m \geq 2$ and $n - m > (k-2)m$ we have $n - m > k - 1$, therefore $n - m > w_1 + w_2 - 1$. By Lemma 6.1.1 there exists an $\mathsf{SHF}(k-2; n_1, m, \{w_1, w_2 - 1\})$ with

$$n_1 \geq n - m > (k-1)m - m = (k-2)m,$$

which contradicts the assumption of the induction. This completes the proof. □

Using Lemma 2.2.11 and Theorem 6.1.2 we obtain a new bound for arbitrary N.

Theorem 6.1.3. *Suppose there exists an* $\mathsf{SHF}(N; n, m, \{w_1, w_2\})$. *Let* $u = w_1 + w_2$. *Then* $n \leq (u-1)m^{\lceil \frac{N}{(u-1)} \rceil}$.

Proof. Assume, by contradiction, that there exists an $\mathsf{SHF}(N; n, m, \{w_1, w_2\})$ with $n = (u-1)m^{\lceil \frac{N}{(u-1)} \rceil} + 1$. By Lemma 2.2.11 there exists an $\mathsf{SHF}(\lceil \frac{N}{c} \rceil; n, m^c, \{w_1, w_2\})$ with $c := \lceil \frac{N}{(u-1)} \rceil$. We make use of a simple observation. Suppose there exists an $\mathsf{SHF}(N; n, m, \{w_1, w_2, \ldots, w_t\})$ with matrix representation $\mathcal{A}$. Then for any $N' > N$ there exists an $\mathsf{SHF}(N'; n, m, \{w_1, w_2, \ldots, w_t\})$ obtained by adding $N' - N$ arbitrary new rows using the same symbol set to $\mathcal{A}$. Now, as $\lceil \frac{N}{c} \rceil \leq u - 1$, the observation says that there is an $\mathsf{SHF}(u-1; n, m^c, \{w_1, w_2\})$ with $n = (u-1)m^{\lceil \frac{N}{(u-1)} \rceil} + 1$, which contradicts Theorem 6.1.2. □

As an immediate consequence of Theorem 6.1.3 we have the following result.

Theorem 6.1.4. *Suppose there exists an* $\mathsf{SHF}(N; n, m, \{w_1, \ldots, w_t\})$ *Let* $u = \sum_{i=1}^{t} w_i$. *Then*

$$n \leq (u-1)m^{\lceil \frac{N}{u-1} \rceil}.$$

Proof. The theorem follows from the observation that if there exists an SHF with parameters $(N; n, m, \{w_1, w_2, \ldots, w_t\})$ where $t \geq 3$, then there exists an $\mathsf{SHF}(N; n, m, \{w_1, w_2'\})$ where $w_2' = w_2 + \cdots + w_t$ (Theorem 2.2.7). □

6.2 Second general bound

Here we prove our second new bound on n in an $\mathsf{SHF}(N; n, m, \{w_1, \ldots, w_t\})$ with $t \geq 3$. We prove that

$$n \leq (u-1)(m^{\lceil \frac{N}{(u-1)} \rceil} - 1) + 1.$$

The bound presented in this section provides a stronger bound on the number of columns of a general separating hash family with $t \geq 3$ than the bound of Section 6.1. For $t = 2$ the bound of Section 6.1 remains the best.

In this section, we modify the technique used to prove Theorem 3.6.4 and obtain a stronger upper bound on n for $t \geq 3$. First we prove the bound for $N = u - 1$ and $t \geq 3$. Then we apply Lemma 2.2.11 to generalize the bound to arbitrary N.

Theorem 6.2.1. *Let $t \geq 3$ be an integer. Suppose there exists an $\mathsf{SHF}(u-1; n, m, \{w_1, \ldots, w_t\})$ where $u = \sum_{i=1}^{t} w_i$ and $w_1 \leq w_i$ for $i = 2, \ldots, t$. Then $n \leq (u-1)(m-1)+1$.*

Proof. Assume, for a contradiction that there exists an $\mathsf{SHF}(u-1; n, m, \{w_1, \ldots, w_t\})$ with $n = (u-1)(m-1)+2$. Wlog we assume that w_1 and w_2 are the smallest two of the integers $w_1, w_2, \ldots, w_t$. Let $\mathcal{A} = (a_{i,j})$ be its matrix representation and $\mathcal{C}$ denote the set of columns of $\mathcal{A}$. The proof describes a procedure how to construct disjoint subsets $\mathsf{C}_1, \mathsf{C}_2, \ldots, \mathsf{C}_t \subseteq \mathcal{C}$ with $|\mathsf{C}_i| \leq w_i$ that are not separated by any row of $\mathcal{A}$. We begin with a simple counting of the number of columns having at least one unique symbol in some row $i \in \{2, \ldots, u-1\}$. Since each row can have at most $m-1$ unique symbols (if there were m unique symbols, we would only have m columns), there are at most $(u-2)(m-1)$ such columns. Let $\mathcal{C}_1$ denote this set of columns. Hence, $|\mathcal{C}_1| \leq (u-2)(m-1)$. Define $\mathcal{C}_2 := \mathcal{C} \setminus \mathcal{C}_1$. Then

$$|\mathcal{C}_2| = |\mathcal{C}| - |\mathcal{C}_1| \geq (u-1)(m-1) + 2 - (u-2)(m-1) \geq m+1.$$

The set $\mathcal{C}_2$ has the following property: For each column $j \in \mathcal{C}_2$ and each row $i \in \{2, \ldots, u-1\}$, the symbol $a_{i,j}$ appears in row i at least two times. As $|\mathcal{C}_2| \geq m+1$, it follows that there are two columns $j_1, j_2 \in \mathcal{C}_2$ having the same symbol in the first row and having non-unique symbols in all other rows.

We now use the repeating property of elements in columns j_1 and j_2 and describe how to construct the subsets $\mathsf{C}_1, \mathsf{C}_2, \ldots, \mathsf{C}_t$ of $\mathcal{C}$ we are seeking. We start with $\mathsf{C}_i = \emptyset$ for $i = 1, \ldots, t$ and then construct C_i's using the following four steps.

Step 1: Add j_1 to C_1 and j_2 to C_2. We will focus on the specified columns j_1 and j_2 in the following steps to construct C_1, C_2, $\ldots$, C_t.

Step 2: This step starts building sets C_i for $i = 3, \ldots, t$ such that they are not separated from column j_1 in rows $k = 2, \ldots, u - w_1 - w_2 + 1$.

Consider all the rows $k = 2, \ldots, u - w_1 - w_2 + 1$ of $\mathcal{A}$. For each such row k, the symbol a_{k,j_1} appears in at least one more column, say j, other than j_1 (i.e. $j \neq j_1$).

(i) If $j \in \bigcup_{i=3}^{t} \mathsf{C}_i \cup \mathsf{C}_2$, then do nothing. As in this case, C_1 is not separated from $\mathsf{C}_2, \ldots, \mathsf{C}_t$ in row k.

(ii) If $j \notin \bigcup_{i=3}^{t} \mathsf{C}_i \cup \mathsf{C}_2$ and if $|\mathsf{C}_i| < w_i$ for some $i = 3, \ldots, t$, then add column j to set C_i.

In this step we can always find some $3 \leq i \leq t$ with $|\mathsf{C}_i| < w_i$, while at the beginning of this step we have $|\mathsf{C}_i| = 0$ for $i = 3, \ldots, t$. For each row $k = 2, \ldots, u - w_1 - w_2 + 1$, we add at most one element to $\bigcup_{i=3}^{t} \mathsf{C}_i$ which leads to at most $u - w_1 - w_2$ elements. At the end of this step, after considering all the rows $k = 2, \ldots, u - w_1 - w_2 + 1$, we will have $|\bigcup_{i=3}^{t} \mathsf{C}_i| \leq u - w_1 - w_2 = \sum_{i=3}^{t} w_i$.

We eventually obtain subsets $\mathsf{C}_3, \ldots, \mathsf{C}_t$ with $|\mathsf{C}_i| \leq w_i$ that together with C_2 are not separated from column j_1 in any row $k = 1, \ldots, u - w_1 - w_2 + 1$. Note that after Step 2 all sets $\mathsf{C}_3, \ldots, \mathsf{C}_t$ could remain empty, this would be the case if column j is unique and $j = j_2$ for all k.

Step 3: This step continues to construct the sets $\mathsf{C}_3, \ldots, \mathsf{C}_t$ as long as it is still possible, otherwise it constructs the set C_1.

Consider all the rows $k = u - w_1 - w_2 + 2, \ldots, u - w_2$ ($w_1 - 1$ rows). In each row k there exists a column j with $j \neq j_2$ such that $a_{k,j} = a_{k,j_2}$ (as the symbol a_{k,j_2} is repeated).

(i) If column $j \in \bigcup_{i=3}^{t} \mathsf{C}_i$, then do nothing.

(ii) If column $j \notin \bigcup_{i=3}^{t} \mathsf{C}_i \cup \mathsf{C}_1$ and if $\sum_{i=3}^{t} |\mathsf{C}_i| < w_3 + \cdots + w_t$ (i.e. there exists some $i = 3, \ldots, t$ with $|\mathsf{C}_i| < w_i$), then add j to one of C_i with $|\mathsf{C}_i| < w_i$, $i = 3, \ldots, t$.

(iii) If column $j \notin \bigcup_{i=3}^{t} \mathsf{C}_i \cup \mathsf{C}_1$ and if $\sum_{i=3}^{t} |\mathsf{C}_i| = w_3 + \cdots + w_t$, then add j to C_1.

(iv) If column $j \in \mathsf{C}_1$, then do nothing.

Note that before Step 3 we have $\mathsf{C}_1 = \{j_1\}$. In Step 3 for each of $w_1 - 1$ considered rows we add at most one column to C_1. So we have $|\mathsf{C}_1| \leq w_1$ after Step 3. It also leads to $|\mathsf{C}_1| < w_1$ in each row $k = u - w_1 - w_2 + 2, \ldots, u - w_2$ and results that we can add new elements to C_1.

This process in Step 3 is characterized by the following property: By finishing Step 3, if $|\mathsf{C}_1| \geq 2$, then $\sum_{i=3}^{t} |\mathsf{C}_i| = w_3 + \cdots + w_t$ (i.e. $|\mathsf{C}_i| = w_i$ for all $i = 3, \ldots, t$).

It is clear that $\mathsf{C}_1, \mathsf{C}_2, \mathsf{C}_3, \ldots, \mathsf{C}_t$ are not separated in any row $k = u - w_1 - w_2 + 2, \ldots, u - w_1$.

Define a set D_2 as follows: D_2 is the set of columns j obtained from (i) and (ii) of Step 3 after it is finished. Note here that $\mathsf{D}_2 \cup \mathsf{C}_1$ is the set of columns that are responsible for the non-separation of C_2 from $\mathsf{C}_1, \mathsf{C}_3, \ldots, \mathsf{C}_t$ in the rows $k = u - w_1 - w_2 + 2, \ldots, u - w_2$. Define $\mathsf{D}_1 := \bigcup_{i=3}^{t} \mathsf{C}_i \setminus \mathsf{D}_2$.

Step 4: This step essentially deals with the extension of C_2 by using rows $k = u - w_2 + 1, \ldots, u - 1$. A crucial point of this step is that we might need to modify the so far constructed sets $\mathsf{C}_1, \mathsf{C}_3, \ldots, \mathsf{C}_t$. To make the description more clear, we consider two cases

Case A: $|\mathsf{C}_1| = 1$ (i.e. $\mathsf{C}_1 = \{j_1\}$).

For each $k = u - w_2 + 1, \ldots, u - 1$, there exists a column $j \neq j_1$ such that $a_{k,j} = a_{k,j_1}$, as the symbol a_{k,j_1} is repeated.

(a) If $j \in \bigcup_{i=3}^{t} \mathsf{C}_i \cup \mathsf{C}_2$, do nothing.

(b) If $j \notin \bigcup_{i=3}^{t} \mathsf{C}_i \cup \mathsf{C}_2$, add j to C_2.

It can be checked that the set C_1 is not separated from the sets $\mathsf{C}_2, \mathsf{C}_3, \ldots, \mathsf{C}_t$ in any row $k = u - w_2 + 1, \ldots, u - 1$. Therefor, the constructed sets $\mathsf{C}_1, \ldots, \mathsf{C}_t$ are not separated in any row $k = 1, \ldots, u - 1$.

Case B: $|\mathsf{C}_1| \geq 2$.

Suppose $|\mathsf{C}_1| := \alpha \geq 2$. As just described in Step 3 this case implies that $|\mathsf{C}_i| = w_i$ for all $i = 3, \ldots, t$. Moreover, we have $\bigcup_{i=3}^{t} \mathsf{C}_i = \mathsf{D}_1 \cup \mathsf{D}_2$ as defined in Step 3.

As mentioned in step 3, columns in $\mathsf{D}_2 \cup \mathsf{C}_1$ agree with C_2 in rows $k = u - w_1 - w_2 + 2, \ldots, u - w_2$, i.e. totally $w_1 - 1$ rows. In each row $k = u - w_1 - w_2 + 2, \ldots, u - w_2$, the symbol in C_2 is the same as the symbol in some column c_k in $\mathsf{D}_2 \cup \mathsf{C}_1$. For $k \neq k'$, we may have $c_k = c_{k'}$ which means that $|\mathsf{D}_2 \cup \mathsf{C}_1| \leq w_1$ (Considering also $j_1 \in \mathsf{C}_1$). Since $\alpha - 1$ columns are added to C_2 in Step 3, we have

$$|\mathsf{D}_2| = w_2 - 1 - (\alpha - 1) = w_2 - \alpha.$$

Further, as

$$w_1 \leq w_3 \leq |\bigcup_{i=3}^{t} \mathsf{C}_i| = w_3 + \ldots + w_t = |\mathsf{D}_1| + |\mathsf{D}_2| = |\mathsf{D}_1| + w_1 - \alpha,$$

we have

$$|\mathsf{D}_1| \geq \alpha.$$

We now use this fact to construct C_2 or possibly to modify the so far constructed $\mathsf{C}_1, \mathsf{C}_3, \ldots, \mathsf{C}_t$.

For each row $k = u - w_2 + 1, \ldots, u - 1$, there exists a column $j \neq j_1$ such that $a_{k,j} = a_{k,j_1}$, as the symbol a_{k,j_1} is repeated.

(i) If $j \in \bigcup_{i=3}^{t} \mathsf{C}_i \cup \mathsf{C}_2$, do nothing.

(ii) If $j \notin \bigcup_{i=3}^{t} \mathsf{C}_i \cup \mathsf{C}_2 \cup \mathsf{C}_1$, add j to C_2.

(iii) If $j \in \mathsf{C}_1$ (i.e. cases (i) and (ii) do not happen), then we do the following operation: Move one column $j' \in \mathsf{D}_1$ to C_2 and substitute j' with j. We observe that this step can always be done, as $|\mathsf{D}_1| \geq \alpha = |\mathsf{C}_1|$. Note that the size of C_1 is reduced by one each time this operation is applied.

Now it is not difficult to check that the constructed column subsets $\mathsf{C}_1, \mathsf{C}_2, \mathsf{C}_3, \ldots, \mathsf{C}_t$ cannot be separated by any row of $\mathcal{A}$. This can be seen as follows. After Steps 1,2,3 the so far constructed $\mathsf{C}_1, \mathsf{C}_2, \mathsf{C}_3, \ldots, \mathsf{C}_t$ are not separated by any of the first $(u - w_2)$ rows of $\mathcal{A}$, (i.e. rows $k = 1, \ldots, u - w_2$). The key observation being that any operation in Step 4, namely adding a new column to C_2 or moving one column from D_1 to C_2 and replacing it by a column from C_1, does not change the non-separation property of the newly constructed sets $\mathsf{C}_1, \mathsf{C}_2, \mathsf{C}_3, \ldots, \mathsf{C}_t$ in rows $k = 1, \ldots, u - w_2$. It can be explained as follows: By adding some column to C_2 with the above conditions, we get non-separation between C_1 and C_2 in some row $k = u - w_2 + 1, \ldots, u - 1$ which does not interfere the non-separation property in rows $k = 1, \ldots, u - w_2$. Furthermore, it is clear from step 3 that the columns in D_1 provide non-separation between C_1 and $\cup_{i=3}^{t} \mathsf{C}_i$ in rows $k = u - w_1 - w_2 + 2, \ldots, u - w_2$. Hence, moving such a column from $\cup_{i=3}^{t} \mathsf{C}_i$ to C_2 reserves non-separation in these rows. Moreover, the construction in Step 4 makes clear that the column sets $\mathsf{C}_1, \mathsf{C}_2, \mathsf{C}_3, \ldots, \mathsf{C}_t$ are not separated by any of the last $(w_2 - 1)$ rows, i.e. rows $k = u - w_2 + 1, \ldots, u - 1$. This completes the proof. □

From Theorem 6.2.1 and Lemma 2.2.11 we obtain the following bound for arbitrary N.

Theorem 6.2.2. *Let $t \geq 3$ be an integer. Suppose there exists an $\mathsf{SHF}(N; n, m, \{w_1, \ldots, w_t\})$ with $w_1 \leq w_i$ for $i = 2, \ldots, t$. Let $u = \sum_{i=1}^{t} w_i$. Then $n \leq (u - 1)(m^{\lceil \frac{N}{u-1} \rceil} - 1) + 1$.*

6.3 Third general bound

In this section, we derive a new necessary condition for the existence of an $\mathsf{SHF}(N; n, m, \{w_1, \ldots, w_t\})$ when $t \geq 3$ and $u \geq 4$. It must be mentioned that in an $\mathsf{SHF}(N; n, m, \{w_1, w_2, w_3\})$ the condition $u \geq 4$ is equivalent to the existence of at least one $i \in \{1, 2, 3\}$, for which $w_i \geq 2$. We assume here that $w_3 \geq 2$. If $t \geq 4$, the condition $u \geq 4$ is sufficient and we can have $w_i = 1$ for $i = 1, \ldots, t$.

Our new bound is a generalization of the bound for an $\mathsf{SHF}(N; n, m, \{1, 1, 2\})$ proved by Stinson et al. in [45] (see section 3.7).

Theorem 6.3.1. *[45] If an* $\mathsf{SHF}(3; n, m, \{1, 1, 2\})$ *exists, then*

$$n \leq 3m + 2 - 2\sqrt{3m + 1}.$$

which results in next bound.

Theorem 6.3.2. *[45] If an* $\mathsf{SHF}(N; n, m, \{1, 1, 2\})$ *exists, then*

$$n \leq 3m^{\lceil \frac{N}{3} \rceil} + 2 - 2\sqrt{3m^{\lceil \frac{N}{3} \rceil} + 1}.$$

The aim is to prove that in an $\mathsf{SHF}(N; n, m, \{w_1, \ldots, w_t\})$ with $t \geq 3$ and $u = \sum_{i=1}^{t} w_i$ if $u \geq 4$ we have

$$n \leq (u-1)m^{\lceil \frac{N}{u-1} \rceil} + 2 - 2\sqrt{3m^{\lceil \frac{N}{u-1} \rceil} + 1}. \tag{6.1}$$

To get this result, we first prove that (6.1) holds in an $\mathsf{SHF}(u-1; n, m, \{w_1, w_2, w_3\})$ (Theorem 6.3.4). Then we generalize it to arbitrary t and finally get the bound for arbitrary N.

We use the generalization of Lemma 6.1.1 which is proved below in the proof of Theorem 6.3.4.

Lemma 6.3.3. *Suppose there exists an* $\mathsf{SHF}(N; n, m, \{w_1, \ldots, w_t\})$ *with* $n-m \geq u-1$ *where* $u = \sum_{i=1}^{t} w_i$ *and* $w_t \geq 2$. *Then there exists an* $\mathsf{SHF}(N-1; n_1, m, \{w_1, \ldots, w_{t-1}, w_t - 1\})$ *with* $n_1 \geq n - m$.

Proof. According to Remark 2.2.1, we have $N - 1 \geq 1$ and $n_1 \geq 1$.

Let $\mathcal{A}$ be the matrix representation of an $\mathsf{SHF}(N; n, m, \{w_1, \ldots, w_t\})$ with $w_t \geq 2$. Let $m_1 \leq m$ denote the number of symbols that appear in the first row of $\mathcal{A}$. Since permuting the columns of $\mathcal{A}$ does not change the separation property (according to Lemma 2.2.4), we may assume that the first row of $\mathcal{A}$ has pairwise different symbols in the first m_1 columns. Let $\mathcal{A}_1$ denote the $(N-1) \times (n-m_1)$ matrix obtained from $\mathcal{A}$ by ignoring the first row and the first m_1 columns of $\mathcal{A}$. Set $n_1 := n - m_1$. Then $n_1 \geq n - m$. We claim that $\mathcal{A}_1$ is an $\mathsf{SHF}(N-1; n_1, m, \{w_1, \ldots, w_{t-1}, w_t - 1\})$.

Assume that $\mathcal{A}_1$ is not an $\mathsf{SHF}(N-1; n_1, m, \{w_1, \ldots, w_{t-1}, w_t - 1\})$. Then there are t column sets $\mathsf{C}_1, \ldots, \mathsf{C}_t$ with $|\mathsf{C}_i| = w_i$ for $i = 1, \ldots, t-1$ and $|\mathsf{C}_t| = w_t - 1$, that are not separated in any row of $\mathcal{A}_1$. Now consider the sets of columns $\mathsf{C}_1, \ldots, \mathsf{C}_t$ in matrix $\mathcal{A}$. Let a be a symbol appearing in some column of C_1 in the first row of $\mathcal{A}$. Then in the first m_1 columns of $\mathcal{A}$ there is a column c having symbol a in the first row. Add this column c to C_t. Now it is easily checked that $\mathsf{C}_1, \ldots, \mathsf{C}_{t-1}$ and $\mathsf{C}_t \cup \{c\}$ are not separated in $\mathcal{A}$, which contradicts the separating property of $\mathcal{A}$. □

Now we can prove the following necessary condition. We assume that in the separating hash families considered in the next theorems we have $n - m \geq u - 1$ so that we can use Lemma 6.3.3.

Theorem 6.3.4. *In an* $\mathsf{SHF}(u-1; n, m, \{w_1, w_2, w_3\})$ *with* $u = w_1 + w_2 + w_3$ *and* $w_3 \geq 2$ *and* $n - m \geq u - 1$*, we have*

$$n \leq (u-1)m + 2 - 2\sqrt{3m+1}.$$

Proof. The theorem is proved by induction on u.

As $t = 3$ and at most two of the w_i's can have the value 1, the smallest possible value for u is 4 and results in type $\{1, 1, 2\}$.

Step 1. Let $u = 4$ and $\mathcal{F}$ be an $\mathsf{SHF}(3; n, m, \{1, 1, 2\})$. Theorem 6.3.1 shows that the theorem is true in this case.

Now assume that for $u = k - 1 \geq 4$ we have

$$n \leq (u-1)m + 2 - 2\sqrt{3m+1}$$

in an $\mathsf{SHF}(u-1; n, m, \{w_1, w_2, w_3\})$ with $u = w_1 + w_2 + w_3$ and $w_3 \geq 2$.

Step 2. Let $u = k$ and $\mathcal{F}$ be an $\mathsf{SHF}(k-1; n, m, \{w_1, w_2, w_3\})$ with $k = w_1 + w_2 + w_3$ and $w_3 \geq 2$ where

$$n = (k-1)m + 2 - 2\sqrt{3m+1} + 1.$$

According to Lemma 6.3.3 there exists an $\mathsf{SHF}(k-2; n_1, m, \{w_1, w_2, w_3 - 1\})$ with

$$n_1 \geq n - m = (k-1)m + 2 - 2\sqrt{3m+1} + 1 - m = (k-2)m + 2 - 2\sqrt{3m+1} + 1.$$

If $w_3 > 2$ then $w_3 - 1 \geq 2$ and the assumption of the induction gives the contradiction.

Let $w_3 = 2$. As $k - 1 \geq 4$ we have $k \geq 5$ which means that

$$w_1 + w_2 + w_3 = w_1 + w_2 + 2 \geq 5 \Rightarrow w_1 + w_2 \geq 3.$$

We can assume that $w_2 \geq 2$ and conclude that there exists an $\mathsf{SHF}(k-2; n_1, m, \{w_1', w_2', w_3'\})$ where $w_1' = w_1$, $w_2' = w_3 - 1$ and $w_3' = w_2$ which contradicts the assumption of the induction. □

Now we prove the result for $t > 3$.

Theorem 6.3.5. *In an* $\mathsf{SHF}(u-1; n, m, \{w_1, \ldots, w_t\})$ *with* $t \geq 3$ *and* $u = \sum_{i=1}^{t} w_i \geq 4$ *we have*

$$n \leq (u-1)m + 2 - 2\sqrt{3m+1}.$$

Proof. Suppose there exists an $\mathsf{SHF}(u-1; n, m, \{w_1, \ldots, w_t\})$ with $t \geq 3$ and $u = \sum_{i=1}^{t} w_i \geq 4$ such that

$$n = (u-1)m + 2 - 2\sqrt{3m+1} + 1.$$

Here we may assume that w_1 and w_2 are the smallest of $w_1, \ldots, w_t$. From Lemma 2.2.7 there exists an $\mathsf{SHF}(u-1; n, m, \{w_1, w_2, w_3'\})$ with $u = w_1 + w_2 + w_3'$ and $w_3' = \sum_{i=3}^{t} w_i \geq 2$ and

$$n = (u-1)m + 2 - 2\sqrt{3m+1} + 1.$$

which is a contradiction to Theorem 6.3.4. □

The result in Theorem 6.3.5 provides a bound for arbitrary N using Lemma 2.2.11.

Theorem 6.3.6. *In an* $\mathsf{SHF}(N; n, m, \{w_1, \ldots, w_t\})$ *with* $u = \sum_{i=1}^{t} w_i \geq 4$ *and* $t \geq 3$ *we have*

$$n \leq (u-1)m^{\lceil \frac{N}{u-1} \rceil} + 2 - 2\sqrt{3m^{\lceil \frac{N}{u-1} \rceil} + 1}.$$

Let $\tilde{u} = \sum_{i=1}^{t} w_i \geq 4$ be a fixed value. For an $\mathsf{SHF}(N; n, m, \{w_1, \ldots, w_t\})$ with $\tilde{u} = \sum_{i=1}^{t} w_i$ and $t \geq 3$ Theorem 6.3.6 provides the following bound:

$$n \leq (\tilde{u}-1)m^{\lceil \frac{N}{\tilde{u}-1} \rceil} + 2 - 2\sqrt{3m^{\lceil \frac{N}{\tilde{u}-1} \rceil} + 1}, \tag{6.2}$$

while Theorem 6.2.2 results in:

$$n \leq (\tilde{u}-1)(m^{\lceil \frac{N}{\tilde{u}-1} \rceil} - 1) + 1. \tag{6.3}$$

It is observed that by increasing m the bound of (6.2) is stronger than the bound in (6.3). However, when m is fixed and the type is changed such that u increases, then Theorem 6.2.2 improves the bound of Theorem 6.3.6.

6.4 Summary

We provided new upper bounds on the number of columns of a separating hash family in this chapter. These new bounds are summarized in Table 6.1.

Table 6.1: New bounds for general SHFs

Theorem	Parameters	Conditions	New bounds
6.1.4	$(N; n, m, \{w_1, \ldots, w_t\})$		$(u-1)m^{\lceil \frac{N}{u-1} \rceil}$
6.2.2	$(N; n, m, \{w_1, \ldots, w_t\})$	$t \geq 3$, $w_1 \leq w_i, i = 2, \ldots, t$	$(u-1)(m^{\lceil \frac{N}{u-1} \rceil} - 1) + 1$
6.3.6	$(N; n, m, \{w_1, \ldots, w_t\})$	$u \geq 4, t \geq 3$	$(u-1)m^{\lceil \frac{N}{u-1} \rceil} + 2 - 2\sqrt{3m^{\lceil \frac{N}{u-1} \rceil} + 1}$

The bound of Theorem 6.1.4 is the only general bound which holds for all parameters without any restriction. When $t \geq 3$, Theorem 6.2.2 provides a stronger bound. In addition to the last condition, if $u \geq 4$ then the bound of Theorem 6.3.6 is stronger than the two previously mentioned bounds.

Table 6.2 gives the result of comparing the bound of Theorem 6.1.4 with the bounds in Chapter 3.

Table 6.2: Comparison of Theorem 6.1.4 with other bounds

Parameters	Conditions	Theorem	Bound
$(N; n, m, \{1, 1\})$		**6.1.4**	m^N
		3.1.1	m^N
$(N; n, m, \{\overbrace{1, 1, \ldots, 1}^{t}\})$	$t \geq 3$	**6.1.4**	$(t-1)m^{\lceil \frac{N}{t-1} \rceil}$
		3.1.2	$(t-1)(m^{\lceil \frac{N}{t-1} \rceil} - 1)$
$(N; n, m, \{1, w\})$	$w \geq 2$	**6.1.4**	$wm^{\lceil \frac{N}{w} \rceil}$
		3.2.1	$w(m^{\lceil \frac{N}{w} \rceil} - 1)$
$(N; n, m, \{1, w\})$	$2 \leq N \leq w$	**6.1.4**	wm
		3.2.2	$N(m-1)$
$(N; n, m, \{w, w\})$		**6.1.4**	$(2w-1)m^{\lceil \frac{N}{2w-1} \rceil}$
		3.3.4	$(2w^2 - 3w + 2)m^{\lceil \frac{N}{2w-1} \rceil} - 2w^2 + 3w - 1$
$(N; n, m, \{w, w-1\})$		**6.1.4**	$(2w-2)m^{\lceil \frac{N}{2w-2} \rceil}$
		3.4.2	$(2w^2 - 5w + 4)m^{\lceil \frac{N}{2w-2} \rceil} - 2w^2 + 5w - 3$
$(N; n, m, \{w_1, \ldots, w_t\})$	$w_1, w_2 \leq w_i,$	**6.1.4**	$(u-1)m^{\lceil \frac{N}{u-1} \rceil}$
	$i = 3, \ldots, t$	3.6.3	$(w_1 w_2 + u - w_1 - w_2)m^{\lceil \frac{N}{u-1} \rceil}$

Chapter 7

Improved bounds on SHFs

In Chapter 6 we provided three different upper bounds on general separating hash families. The interesting question here is if these bounds can be improved or they are strong bounds in all cases. The aim of this chapter is to give some improved bounds. We consider special cases and prove bounds which are stronger than the previous general bounds. At the end of the chapter, we compare all the bounds to observe which bound gives the lowest value in different cases.

7.1 $\mathsf{SHF}(u-2; n, m, \{w_1, w_2\})$

In this section we consider separating hash families of type $\{w_1, w_2\}$ with $N = w_1 + w_2 - 2$ rows. We also assume that $w_1 \leq w_2$ and $w_2 \geq 3$. We prove in an $\mathsf{SHF}(w_1 + w_2 - 2; n, m, \{w_1, w_2\})$ the number of columns is bounded by:

$$n \leq (w_2 - 1)(m - 1) + w_2. \tag{7.1}$$

Theorem 6.1.4 gives the following bound on an $\mathsf{SHF}(w_1 + w_2 - 2; n, m, \{w_1, w_2\})$:

$$n \leq (w_1 + w_2 - 1)m. \tag{7.2}$$

Note that the bound in (7.1) is better than the bound in (7.2) as shown below.

As $m \geq 2$ we have:

$$\begin{aligned} (w_2 - 1)(m - 1) + w_2 &< (w_2 - 1)(m - 1) + w_2 + w_1 m \\ &= (w_1 + w_2 - 1)m + 1 \end{aligned} \tag{7.3}$$

Hence:

$$(w_2 - 1)(m - 1) + w_2 \leq (w_1 + w_2 - 1)m.$$

In the following lemma, we prove our bound.

Lemma 7.1.1. *If there is an* $\mathsf{SHF}(w_1 + w_2 - 2; n, m, \{w_1, w_2\})$ *with* $w_1 \leq w_2$ *and* $w_2 \geq 3$*, then* $n \leq (w_2 - 1)(m - 1) + w_2$.

Proof. Assume that $\mathcal{A}$ is the matrix representation of an $\mathsf{SHF}(w_1 + w_2 - 2; (w_2 - 1)(m-1) + w_2 + 1, m, \{w_1, w_2\})$ where $w_1 \leq w_2$ and $w_2 \geq 3$. As $n > (w_2 - 1)(m-1)$ and $w_2 - 1, m \geq 2$ according to Lemma 4.1.1 there is some column with all repeating elements in the first $w_2 - 1$ rows as shown below:

a_1	a_1	$*$	$*$	$*$
a_2	$*$	a_2	$*$	$*$
$\vdots$	$*$	$*$	$\ddots$	$*$
a_{w_2-1}	$*$	$*$	$*$	a_{w_2-1}

(7.4)

Let $\mathcal{C}$ denote the set of columns of $\mathcal{A}$ and $\mathcal{C}_1$ be the set of columns of the array (7.4). Let $\mathcal{C}_2 = \mathcal{C} \setminus \mathcal{C}_1$. Then

$$|\mathcal{C}_2| = (w_2 - 1)(m-1) + w_2 + 1 - w_2 > (w_2 - 1)(m-1) \geq (w_1 - 1)(m-1).$$

So according to Lemma 4.1.1, in the last $w_1 - 1$ rows there is a column with all repeating elements. The following array shows the situation:

a_1	a_1	$*$	$*$	$*$	$*$	$*$	$*$	$*$	$*$
a_2	$*$	a_2	$*$	$*$	$*$	$*$	$*$	$*$	$*$
$\vdots$	$*$	$*$	$\ddots$	$*$	$*$	$*$	$*$	$*$	$*$
a_{w_2-1}	$*$	$*$	$*$	a_{w_2-1}	$*$	$*$	$*$	$*$	$*$
$*$	$*$	$*$	$*$	$*$	b_1	b_1	$*$	$*$	$*$
$*$	$*$	$*$	$*$	$*$	b_2	$*$	b_2	$*$	$*$
$*$	$*$	$*$	$*$	$*$	$\vdots$	$*$	$*$	$\ddots$	$*$
$*$	$*$	$*$	$*$	$*$	b_{w_1-1}	$*$	$*$	$*$	b_{w_1-1}

It is easy to check that the sets $\mathsf{C}_1 = \{2, 3, \ldots, w_2, w_2{+}1\}$ and $\mathsf{C}_2 = \{1, w_2{+}2, \ldots, w_1{+}w_2\}$ are not separated. □

Remark 7.1.2. *It may happen that the repeating elements $a_1, a_2, \ldots, a_{w_2-1}$ do not appear in separated columns like above. Some of them can appear in the same column. The same case can be considered for $b_1, b_2, \ldots, b_{w_1-1}$. In this case we will neither have a separating hash family of type $\{p, q\}$ in which $p \leq w_2$ and $q \leq w_1$ nor a separating hash family of type $\{w_1, w_2\}$.*

By applying Lemma 2.2.11 and Lemma 7.1.1 we obtain a bound for an $\mathsf{SHF}(N; n, m, \{w_1, w_2\})$ with arbitrary N.

Theorem 7.1.3. *If there is an $\mathsf{SHF}(N; n, m, \{w_1, w_2\})$ with $w_1 \leq w_2$ and $w_2 \geq 3$, then $n \leq (w_2 - 1)(m^{\lceil \frac{N}{u-2} \rceil} - 1) + w_2$, where $u = w_1 + w_2$.*

The bound presented in theorem 7.1.3 is an improvement of the bound obtained from 6.1.4 for an $SHF(N; n, m, \{w_1, w_2\})$ when N, w_1 and w_2 satisfy in the following condition:

$$\lceil \frac{N}{w_1 + w_2 - 1} \rceil = \lceil \frac{N}{w_1 + w_2 - 2} \rceil$$

In particular, we get a better bound from 7.1.3 when:

$$w_1 + w_2 - 2 \geq \lceil \frac{N}{2} \rceil.$$

7.2 $\mathsf{SHF}(u; n, m, \{1, w\})$

In this section we investigate separating hash families with parameters $(u; n, m, \{1, w\})$ where $u = 1 + w$ and present a bound on the number of columns in this case [8].

The following lemma is used in the proof of our main result.

Lemma 7.2.1. *Assume that $\mathcal{A} = (a_{i,j}), 1 \leq i \leq N, 1 \leq j \leq n$, is the matrix representation of an $\mathsf{SHF}(N; n, m, \{1, w\}), w < N$. If c_1 and $c_2, 1 \leq c_1, c_2 \leq n$ are two distinct columns of $\mathcal{A}$ which agree in at least $N - (w - 1)$ positions, then there are i_1 and i_2, $1 \leq i_1, i_2 \leq N$, such that a_{i_l,c_l} appears only once in row i_l for $l = 1, 2$.*

Proof. By Lemma 2.2.4, we can permute the rows of $\mathcal{A}$ and assume that the two columns c_1 and c_2 agree in the first l positions, $l \geq N - (w - 1)$. If for each $i = l + 1, \ldots, N$, there is c_i, $1 \leq c_i \neq c_1 \leq n$, such that $a_{i,c_1} = a_{i,c_i}$, then the two sets of columns $\mathsf{C}_1 = \{c_1\}$ and $\mathsf{C}_2 = \{c_2, c_{l+1}, \ldots, c_N\}$, $(|\mathsf{C}_2| \leq w)$, cannot be separated. □

In the following theorem, we present a bound on the number of columns of an $\mathsf{SHF}(u; n, m, \{1, w\})$ where $u = 1 + w$.

Theorem 7.2.2. *In an $\mathsf{SHF}(u; n, m, \{1, w\})$ with $u = 1 + w$ we have*

(i) $n \leq m^2$, *if* $u \leq m$,

(ii) $n \leq um$, *if* $u > m$.

Proof. Assume that $\mathcal{A}$ is the matrix representation of an $\mathsf{SHF}(u; n, m, \{1, w\})$. Let $\mathcal{C}$ denote the set of columns of $\mathcal{A}$. Divide $\mathcal{C}$ into two different parts $\mathcal{C}_1$ and $\mathcal{C}_2$, where $\mathcal{C}_1$ consists of the columns which have the Hamming distance at least $u - 1$ to all other columns and $\mathcal{C}_2 = \mathcal{C} - \mathcal{C}_1$. So, if $c_1 \in \mathcal{C}_2$, then there is at least one column c_2 such that c_1 and c_2 agree in at least two positions and also $c_2 \in \mathcal{C}_2$. According to Lemma 7.2.1, if $c \in \mathcal{C}_2$, then there is some symbol in some row of c which cannot appear anywhere else in that row. For each column c in part $\mathcal{C}_2$ consider one of these symbols and denote it by a_c. Then construct pairwise disjoint sets of columns $\mathsf{C}_1, \mathsf{C}_2, \ldots, \mathsf{C}_u$, each C_i consisting of the columns c having a_c in row i, $1 \leq i \leq u$. If $|\mathsf{C}_i| = l_i, 1 \leq i \leq u$, then $|\mathcal{C}_2| = l_1 + l_2 + \cdots + l_u, 0 \leq l_i \leq m$. On the other hand, $|\mathsf{C}_i| = l_i$ means that $\mathcal{C}_1$ can have at most $m - l_i$ symbols in row $i, 1 \leq i \leq u$. Let $l_{max} = max\{l_i : 1 \leq i \leq u\}$ and $l_{min} = min\{l_i : 1 \leq i \leq u\}$. So there is some $i, 1 \leq i \leq u$ where $l_{min} = l_i$. Thus

in $\mathcal{C}_1$, the i'th row has the maximum possible number of symbols and each symbol can appear at most $m - l_{max}$ times, otherwise two columns in $\mathcal{C}_1$ agree in more than one position. So we have:

$$\begin{aligned} n &= |\mathcal{C}_1| + |\mathcal{C}_2| \leq (m - l_{min})(m - l_{max}) + l_1 + \cdots + l_u \\ &\leq m^2 - m.l_{min} - m.l_{max} + l_{min}.l_{max} + u.l_{max}. \end{aligned}$$

As $l_{max} \leq m$ we will have $l_{max}.l_{min} - m.l_{min} \leq 0$ and so

$$n \leq m^2 + (u - m)l_{max},$$

which is less than or equal to um when $u > m$ and is bounded by m^2 when $u \leq m$. □

Now, if we apply Lemma 2.2.11 to the case $u = w + 1 \geq m$ of Theorem 7.2.2 we obtain the following theorem:

Theorem 7.2.3. *If $N \geq w + 1 \geq m$ then for any $\mathsf{SHF}(N; n, m, \{1, w\})$ we have*

$$n \leq (w + 1)m^{\lceil N/(w+1) \rceil} \tag{7.5}$$

We can easily observe that the bound presented in 7.2.2 is an improvement of the bound obtained from 6.1.4 for an $\mathsf{SHF}(N = u; n, m, \{1, w\})$:

From 6.1.4 we have:

$$n \leq (N - 1)m^2.$$

Now we have one of the following two cases:

(i) $N > m$: As $m \geq 2$ (from 2.2.2) we have $N > 2$. Hence,

$$\begin{aligned} N - 2 > 0 &\Rightarrow 2N - 2 > N \\ &\Rightarrow \tfrac{N}{N-1} < 2 \leq m \\ &\Rightarrow Nm < (N - 1)m^2 \end{aligned}$$

(ii) $N \leq m$: From 2.2.1 ($N \geq 2$), it is obvious that $m^2 \leq (N - 1)m^2$.

For an arbitrary integer m we provide a further direct construction of optimal SHF for Theorem 7.2.2 from mutually orthogonal Latin squares (MOLS). A *Latin square of order m* is an $m \times m$ array consisting of elements of an $m-$set, say S, with the property that each row and each column of the array is a permutation of S. Two $m \times m$ Latin squares are *orthogonal* if no ordered pair occurs more than once when they are superimposed. A set of $t \geq 2$ Latin squares is said to be *mutually orthogonal*, or a set of MOLS, if any two of t squares are orthogonal. Let $\{L_i : 1 \leq i \leq s\}$ be a set of s MOLS on symbols $\{0, 1, \ldots, m - 1\}$. Form an $(s + 2) \times m^2$ array $\mathcal{A} = (a_{ij})$ whose columns are $(i, j, L_1(i, j), L_2(i, j), \ldots, L_s(i, j))^T$ for $0 \leq i, j < m$. Then $\mathcal{A}$ is an orthogonal array, $\mathsf{OA}(2, s + 2, m)$. Now any two columns of $\mathcal{A}$ agree in at most one row, therefore $\mathcal{A}$ forms an $\mathsf{SHF}(s + 2; m^2, m, \{1, s + 1\})$ which is optimal by Theorem 7.2.2 when $s + 2 \leq m$. We have the following.

Theorem 7.2.4. *Suppose there are $w-1$ MOLS of order m with $w+1 \leq m$. Then there exists an optimal $\mathsf{SHF}(w+1; m^2, m, \{1, w\})$.*

Examples 7.2.5. *We consider several small values for m that are not prime powers. It is well-known that there are at least two MOLS of order 10, five MOLS of order 12, three MOLS of order 14 and four MOLS of order 15, see for instance [24]. Hence Theorem 7.2.4 provides the following optimal separating hash families:*
$\mathsf{SHF}(w+1; 10^2, 10, \{1, w\})$ *for* $w = 2, 3$,
$\mathsf{SHF}(w+1; 12^2, 12, \{1, w\})$ *for* $w = 2, 3, 4, 5, 6$,
$\mathsf{SHF}(w+1; 14^2, 14, \{1, w\})$ *for* $w = 2, 3, 4$,
$\mathsf{SHF}(w+1; 15^2, 15, \{1, w\})$ *for* $w = 2, 3, 4, 5$.

By [24, Table 3.81, page 175] it is known that there are at least 6 MOLS of order m for all $m \geq 75$. Thus we have the following theorem.

Theorem 7.2.6. *For any integer $m \geq 75$ there is an optimal $\mathsf{SHF}(w+1; m^2, m, \{1, w\})$ for $w = 2, 3, 4, 5, 6, 7$.*

7.3 $\mathsf{SHF}(u; n, m, \{w_1, w_2, \ldots, w_t\})$

In this section we consider separating hash families satisfying the following two conditions:

(i) The number of rows is $u = w_1 + \cdots + w_t$.

(ii) The number of symbols $m \geq u$.

We prove that in this case the number of columns cannot be greater than m^2. This is clearly an improvement of the bound obtained from Theorem 6.2.2. If we imply Theorem 6.2.2, with the above conditions the obtained bound would be

$$n \leq (u-1)(m^2-1)+1$$

.

First we consider the case $t = 2$ and show that in this case if $w_1 = w_2 \neq 1$ we have $n < m^2$ (In Example 7.3.2, we show that the result does not satisfy for $w_1 = w_2 = 1$). Then in the case $w_1 \neq w_2$ we prove that $n \leq m^2$. This result can be generalized to $t \geq 3$.

It is proved in Lemma 7.3.1 that if there exists an $\mathsf{SHF}(4; n, m, \{2, 2\})$ with $m \geq 4$, then we can not have $n \geq m^2$. It shows that the exponent $\lceil \frac{N}{u-1} \rceil$ in theorem 6.1.4 is not tight in all cases. This lemma is used as the base of the induction in Theorem 7.3.5. We also show in Examples 7.3.3 and 7.3.4 that the condition $m \geq 4$ is necessary for Lemma 7.3.1.

Lemma 7.3.1. *In an $\mathsf{SHF}(4; n, m, \{2, 2\})$ with $m \geq 4$ we have $n < m^2$.*

Proof. Assume that there is an $\mathsf{SHF}(4; m^2, m, \{2,2\})$ with $m \geq 4$, namely $\mathcal{F}$. We prove this lemma in two steps:

(i) First we prove that $\mathcal{F}$ satisfies in the following two conditions:

  (a) Every pair appears in every two rows exactly one time.

  (b) There exist two columns which have different symbols in every row.

(ii) Finally we conclude that a $4 \times m^2$ array with conditions in (a) and (b) cannot be separating of type $\{2,2\}$.

Now we prove the statements (i)-(ii).

(i-a) The number of possible pairs on m elements is m^2. So in each two rows, the number of pairs which can be used to fill the columns of the array is equal to the number of columns. First we show that in any two rows it is not possible that one pair appears in two columns. It means that every two columns of an $\mathsf{SHF}(4; m^2, m, \{2,2\})$ can agree in at most one position.

We assume that some pair appears in the first two rows two times. If there is also some pair repeating in the second two rows, then we have the following forbidden configuration in which the sets of columns $\{1,3\}$ and $\{2,4\}$ are not separable:

a	a	$*$	$*$
b	b	$*$	$*$
$*$	$*$	c	d
$*$	$*$	c	d

It implies that in the last two rows every pair appears exactly one time which means each symbol in the last two rows appears exactly $m \geq 4$ times. Assume that the first two columns have the following form:

a	a
b	b
x	z
y	t

As x and t appear in the corresponding rows at least four times, we get the following forbidden configuration in which the sets of columns $\{1,4\}$ and $\{2,3\}$ are not separable:

a	a	$*$	$*$
b	b	$*$	$*$
x	z	x	$*$
y	t	$*$	t

This argumentation shows that in every two rows, each pair appears exactly one time which means that every two columns agree in at most one position

and every element appears in each row exactly m times.

(i-b) Now we use the above result to prove that in an $\mathsf{SHF}(4; m^2, m, \{2,2\})$ with $m \geq 4$ there exist two columns which are different in all four rows.

If it is not the case, then every two columns agree in at least one position. We consider a submatrix consisting of four columns which are equal in the first row (as $m \geq 4$, from (i-a) we conclude that such columns exist) together with one column which is different from them in the first row. So we have the following configuration:

a	a	a	a	b
$*$	$*$	$*$	$*$	$*$
$*$	$*$	$*$	$*$	$*$
$*$	$*$	$*$	$*$	$*$

in which $b \neq a$. Each of the first four columns should agree with the last column at least in one row from the last three rows. According to the Pigeonhole Principle, two columns from the first four columns agree with the last column in the same row. (If $m < 4$, this condition does not satisfy). It implies that two columns from the first four columns agree at least in two positions which contradicts (i-a).

(ii) Knowing the above results, we can easily show that a $4 \times m^2$-array on $m \geq 4$ elements is not separating of type $\{2,2\}$.

Assume that an $\mathsf{SHF}(4; m^2, m, \{2,2\})$ with $m \geq 4$ exists. It consists of two columns which are different in four rows, as shown below:

x	α
y	β
z	γ
t	δ

As every possible pair appears in every two rows exactly one time and each two columns agree at most in one position, we have the following submatrix:

x	α	x	$*$
y	β	β	$*$
z	γ	$*$	z
t	δ	$*$	δ

in which the sets of columns $\{1,2\}$ and $\{3,4\}$ are not separable. □

The following example shows that Lemma 7.3.1 does not satisfy for $w_1 = w_2 = 1$.

Example 7.3.2. A $2 \times m^2$-array consisting of all m^2 pairs on m symbols as its columns is an $\mathsf{SHF}(2; m^2, m, \{1, 1\})$.

m columns				m columns							m columns			
1	1	...	1	2	2	...	2	...	...	...	m	m	...	m
1	2	...	m	1	2	...	m	...	...	...	1	2	...	m

In the following two examples, we see that the condition $m \geq 4$ is necessary in Lemma 7.3.1.

Example 7.3.3. In this example we present an $\mathsf{SHF}(4; 9, 3, \{2, 2\})$:

1	1	1	2	2	2	3	3	3
1	2	3	1	2	3	1	2	3
1	2	3	2	3	1	3	1	2
1	2	3	3	1	2	2	3	1

It is easily observed that in the above array every two columns agree in exactly one position and each symbol appears in each row exactly three times. These two properties help us to prove that the array is separating of type $\{2, 2\}$.

Let $\mathsf{C}_1 = \{c_1, c_2\}$ and $\mathsf{C}_2 = \{c_3, c_4\}$ be two disjoint sets of columns of the array. Using the fact that every two columns have the same symbol exactly in one row, we may assume that c_1 and c_2 have the same symbol, namely x in row one. If c_3 and c_4 have symbols different from x in the first row, then this row separates C_1 from C_2. Otherwise, x can not appear in row one of both c_3 and c_4, as each element appears in each row exactly three times. So we can assume that c_3 and c_4 have the same symbol in row two and get the following configuration:

x	x	x	y
a	b	c	c
$*$	$*$	$*$	$*$
$*$	$*$	$*$	$*$

where $x \neq y$. If $a = c$ or $b = c$, then the pair (x, c) appears in the first two rows two times which contradicts the property of the array. Therefore, row two separates C_1 from C_2.

Moreover, this example shows a $2 - \mathsf{SHF}(4; 9, 3, \{1, 2\})$. It means that for every two sets of columns C_1 and C_2 with $|\mathsf{C}_1| = 1$ and $|\mathsf{C}_2| = 2$ there are at least two rows that separate C_1 from C_2. Let $\mathsf{C}_1 = \{c_1\}$ and $\mathsf{C}_2 = \{c_2, c_3\}$ be two disjoint sets of columns. c_1 has in one row the same symbol as c_3 and in one row the same symbol as c_4. Hence, at least in two rows C_1 and C_2 have different symbols.

The example below shows that if $m = 2$ then Lemma 7.3.1 does not hold.

Example 7.3.4. An $\mathsf{SHF}(4; 4, 2, \{2, 2\})$:

1	1	2	2
1	2	1	2
1	2	2	1
1	1	1	2

In the following, we show that there is no $\mathsf{SHF}(2w; m^2, m, \{w, w\})$ when $m \geq 2w$ and $w \geq 2$.

Theorem 7.3.5. *Assume that $m \geq 2w$ and $w \geq 2$. In an $\mathsf{SHF}(2w; n, m, \{w, w\})$ we have $n < m^2$.*

Proof. We use induction on w to prove this theorem.

(i) We proved in Theorem 7.3.1 that for $w = 2$ this theorem satisfies.

(ii) Assume that there is no $\mathsf{SHF}(2(w-1); m^2, m, \{w-1, w-1\})$.

Now consider a $2w \times m^2$-array. We remove the first two rows. The remaining, is an array with $2(w-1)$ rows and m^2 columns. According to the assumption of the induction, this array is not a separating hash family of type $\{w-1, w-1\}$. So there are two sets of columns C_1 and C_2 of cardinality $w-1$ which are not separable. We add the first two rows to the array. Assume that A_1 is the set consisting of all elements used in the first row of the columns belonging to C_1 and A_2 is the set of all elements used in the second row of the columns belonging to C_2. Let $\mathcal{C}$ denote the set of all columns and $\mathcal{C}' = \mathcal{C} \backslash (\mathsf{C}_1 \cup \mathsf{C}_2)$. Now we consider the first two rows of the columns in $\mathcal{C}'$. Two cases can occur:

- There exist two different columns c_1 and c_2 in $\mathcal{C}'$ such that the element in the second row of c_1 is a member of A_2 and the element in the first row of c_2 is a member of A_1. In this case we see that $\mathsf{C}_1 \cup \{c_1\}$ is not separated from $\mathsf{C}_2 \cup \{c_2\}$, hence the array is not separating of type $\{2, 2\}$.
- Either c_1 (or c_2) does not exist or there exists only one column in which the element in the first row belongs to A_1 and the element in the second row belongs to A_2. It means that in $\mathcal{C}'$ there exist at most $m-1$ elements to fill the first row of at least $m^2 - [2(w-1)+1]$ columns. As $m > 2w$ there is an element which appears $m+1$ times in the first row. So there are two columns c_1 and c_2 in $\mathcal{C}'$ which agree in the first two rows. It means that $\mathsf{C}_1 \cup \{c_1\}$ is not separable from $\mathsf{C}_2 \cup \{c_2\}$. □

The bound of Theorem 7.3.5 is an strong improvement of all the other bounds, as it shows that the exponent is less than 2. In all the bounds proved until now, m has exponent two when $N = u$.
In 7.3.6, we generalize the result of 7.3.5 to arbitrary w_1 and w_2.

Theorem 7.3.6. *Assume that $m \geq w_1 + w_2$. In an $\mathsf{SHF}(w_1 + w_2; n, m, \{w_1, w_2\})$ we have $n \leq m^2$.*

Proof. We prove this theorem using induction on $u = w_1 + w_2$. As $w_1, w_2 \geq 1$, we have $u \geq 2$.

(i) Assume that $u = 2$. So we have $\{w_1, w_2\} = \{1, 1\}$. As $m \geq w_1 + w_2 = 2$, according to Theorem 7.2.2, in an $\mathsf{SHF}(2; n, m, \{1, 1\})$ we have $n \leq m^2$.

(ii) Now assume that in an $\mathsf{SHF}(N = w_1 + w_2; n, m, \{w_1, w_2\})$ with $N = 1, \ldots, u - 1$ we have $n \leq m^2$. Consider that there exists an $\mathsf{SHF}(u = w_1 + w_2; m^2 + 1, m, \{w_1, w_2\})$ with matrix representation $\mathcal{A}$. If for some $i = 1, 2$, $w_i = 1$ then it is known from Theorem 7.2.2 that in an $\mathsf{SHF}(w; n, m, \{1, w - 1\}$ with $w \leq m$ we have $n \leq m^2$. So we can assume that $w_1, w_2 \geq 2$. We delete the first two rows and get a $(w_1 + w_2 - 2) \times (m^2 + 1)$ array which according to the assumption of the induction is not a separating hash family of type $\{w_1 - 1, w_2 - 1\}$. Hence, there exist two sets of columns C_1 and C_2 with $|\mathsf{C}_1| = w_1 - 1$ and $|\mathsf{C}_2| = w_2 - 2$ that are not separated by any row $i = 3, \ldots, u$ in $\mathcal{A}$. Let A_1 denote the set of symbols in the first row of C_1 and A_2 be the set of symbols in the second row of C_2. Let $\mathcal{C}$ denote the set of all columns and $\mathcal{C}' = \mathcal{C} \backslash (\mathsf{C}_1 \cup \mathsf{C}_2)$. Now we consider the first two rows of the columns in $\mathcal{C}'$. Two cases can occur:

- There exist two different columns c_1 and c_2 such that the element in the second row of c_1 is a member of A_2 and the element in the first row of c_2 is a member of A_1. In this case we see that $\mathsf{C}_1 \cup \{c_1\}$ is not separated from $\mathsf{C}_2 \cup \{c_2\}$, hence the array $\mathcal{A}$ is not separating of type $\{w_1, w_2\}$.
- Either c_1 (or c_2) does not exist or there exists only one column in which the element in the first row belongs to A_1 and the element in the second row belongs to A_2. It means that in $\mathcal{C}'$ there exist at most $m - 1$ elements to fill the first row of at least $m^2 - [(w_1 - 1) + (w_2 - 1) + 1]$ columns. As $m > w_1 + w_2$ there is an element which appears $m + 1$ times in the first row. So there are two columns c_1 and c_2 in $\mathcal{C}'$ which agree in the first two rows. It means that $\mathsf{C}_1 \cup \{c_1\}$ is not separable from $\mathsf{C}_2 \cup \{c_2\}$. □

We can generalize Theorem 7.3.6 to $t \geq 3$.

Theorem 7.3.7. *Assume that $m \geq u = \sum_{i=1}^{t} w_i$. In an $\mathsf{SHF}(u; n, m, \{w_1, w_2, \ldots, w_t\})$ we have $n \leq m^2$.*

Proof. If there exists an $\mathsf{SHF}(u; n, m, \{w_1, w_2, \ldots, w_t\})$ with $n > m^2$, then according to Theorem 2.2.7 there exists an $\mathsf{SHF}(u; n, m, \{w_1, w_2 + \cdots + w_t\})$ with $n > m^2$ which contradicts Theorem 7.3.6. □

The results in this section improve the general bounds in Theorems 6.1.4 and 6.2.2 in the case $N = \sum_{i=1}^{t} w_i$. It is observed that in the case $w_i = w_j$ for $1 \leq i \neq j \leq t$ we improve the exponent and in the case $w_i \neq w_j$ for some $1 \leq i \neq j \leq t$ we only improve the coefficient.

7.4 Summary

We tried to improve the general upper bounds in this chapter. These new improved bounds are presented in Table 7.1.

Table 7.1: Improved bounds for general SHFs

Theorem	Parameters	Conditions	New bounds
7.1.3	$(N; n, m, \{w_1, w_2\})$	$w_1 \leq w_2, w_2 \geq 3$	$(w_2 - 1)(m^{\lceil \frac{N}{u-2} \rceil} - 1) + w_2$
7.2.3	$(N; n, m, \{1, w\})$	$N \geq w + 1 \geq m$	$(w+1)m^{\lceil N/(w+1) \rceil}$
7.3.5	$(2w; n, m, \{w, w\})$	$m \geq 2w, w \geq 2$	$m^2 - 1$
7.3.7	$(u; n, m, \{w_1, w_2, \ldots, w_t\})$	$m \geq u$	m^2

Theorem 7.1.3 provides a bound in the case $t = 2$. Hence it can be compared to the bound of Theorem 6.1.4 among the three bounds presented in Chapter 6. As the exponent in 7.1.3 is larger than the exponent in 6.1.4, the bound is in general weaker. However, as explained in Section 7.1 there are parameters for which the bound of Theorem 7.1.3 is stronger. Table 7.2 shows some examples.

Table 7.2: Comparison of Theorem 7.1.3 with other bounds

Parameters	Theorem	obtained bound
$(10; n, 3, \{3, 4\})$	**7.1.3**	28
	6.1.4	54
$(25; n, 5, \{5, 6\})$	**7.1.3**	626
	6.1.4	1250

The comparison of Theorem 7.2.3 with Theorems 6.1.4 and 3.2.1 results that the exponent of Theorem 7.2.3 is smaller, hence a better bound is obtained (Table 7.3).

It is easy to observe that Theorems 7.3.5 and 7.3.7 provide the best upper bounds for the considered parameters.

Table 7.3: Comparison of Theorem 7.2.3 with other bounds

Parameters	Cnditions	Theorem	obtained bound
$(N; n, m, \{1, w\})$	$N \geq w + 1 \geq m$	**7.2.3**	$(w+1)m^{\lceil N/(w+1)\rceil}$
		6.1.4	$wm^{\lceil \frac{N}{w} \rceil}$
		3.2.1	$w(m^{\lceil \frac{N}{w} \rceil} - 1)$

Chapter 8

SHF Constructions

In this chapter we present some results on direct and recursive constructions of separating hash families. One of the motivations for studying these constructions is that they can be used to construct:

- cover free families ([32], Theorem 2.5),
- sandwich free families ([44], Theorem 4.7),
- secure frameproof codes ([44], Theorems 4.9 and 4.10).

In 8.1 we show some of the known constructions for SHFs. In 8.2 some new recursive constructions are presented. Section 8.3 gives a proof for the optimality of two previously known constructions.

8.1 Known constructions

We begin with presenting known constructions for SHFs of type $\{1, 2\}$. In the first construction we have $N = 2$.

Construction 8.1.1. [32] The following array is an optimal $\mathsf{SHF}(2; 2m-2, m, \{1, 2\})$:

$$\begin{pmatrix} 1 & 2 & \dots & m-1 & m & m & \dots & m \\ m & m & \dots & m & 1 & 2 & \dots & m-1 \end{pmatrix}$$

For $N = 3$ we have the following construction.

Construction 8.1.2. [32] The array:

$$\begin{pmatrix} 1 & 1 & \dots & 1 & 1 & 2 & 2 & \dots & 2 & 2 & 2 & \dots & m & m & \dots & m \\ 1 & 2 & \dots & m-1 & m & 1 & 2 & \dots & m-2 & m-1 & m & \dots & 1 & 2 & \dots & m \\ 2 & 3 & \dots & m & 1 & 3 & 4 & \dots & m & 1 & 2 & \dots & 1 & 2 & \dots & m \end{pmatrix}$$

is an optimal $\mathsf{SHF}(3; m^2, m, \{1, 2\})$.

The above construction can be described in the following way:

Let $X = \mathbb{Z}_m$ and index the columns of the array $\mathcal{A} = (a_{i,j}) = (\phi_i(x, y))$ with elements of the set $Y = X \times X$. Then define the elements of $\mathcal{A}$ as follows:

$$\begin{aligned} \phi_1(x, y) &= x \\ \phi_2(x, y) &= y \\ \phi_3(x, y) &= x + y \pmod m \end{aligned}$$

In Theorem 8.1.5, another way for constructing an optimal $\mathsf{SHF}(3; m^2, m, \{1, 2\})$ for $m \equiv 1, 3 \ (mod\ 6)$ is presented, in which Steiner triple systems are used. We begin with a lemma which is used to show that the construction of Theorem 8.1.5 is an $\mathsf{SHF}(3; m^2, m, \{1, 2\})$. Then, we give the definition of Steiner triple systems and finally present the construction.

The following lemma gives a sufficient condition on an SHF of type $\{1, 2\}$.

Lemma 8.1.3. *[32] An $N \times n$ array $\mathcal{A} = (a_{i,j})$ ($N \geq 3$) with entries from $\mathbb{Z}_m$ having the property that any two columns agree in at most one row is an $\mathsf{SHF}(N; n, m, \{1, 2\})$.*

It is followed that Construction 8.1.2 is separating of type $\{1, 2\}$.

Now we give the definition of Steiner triple systems.

Definition 8.1.4. A *Steiner triple system*, $S(2, 3, v)$, is a pair $(\mathbf{V}, \mathcal{B})$ in which:

- $\mathbf{V}$ is a finite set of *points* with $|\mathbf{V}| = v$,
- $\mathcal{B}$ is a family of subsets of $\mathbf{V}$, called *blocks*,
- $|b| = 3$ for every $B \in \mathcal{B}$,
- Every 2-subset of $\mathbf{V}$ is contained in exactly one block.

The following theorem shows how we can construct an optimal $\mathsf{SHF}(3; m^2, m, \{1, 2\})$ using Steiner triple systems of order m, $m \equiv 1, 3 \ (mod\ 6)$.

Construction 8.1.5. [32] An optimal $\mathsf{SHF}(3; m^2, m, \{1, 2\})$ can be constructed using Steiner triple systems of order m, where $m \equiv 1, 3 \ (mod\ 6)$.

Proof. Suppose $m \equiv 1, 3 \ (mod\ 6)$ and let $\mathcal{B}$ be the set consisting of $m(m-1)/6$ triples of a Steiner triple system of order m with point set $\mathbb{Z}_m$. For each $B \in \mathcal{B}$, consider each permutation of the elements of B and make it a column of an array $\mathcal{A}$. Clearly, the array $\mathcal{A}$ contains

$$(3!)m(m-1)/6 = m(m-1) = m^2 - m$$

columns. Because each column of $\mathcal{A}$ is a permutation of the elements in a block of the Steiner triple system, any two columns of $\mathcal{A}$ agree in at most one row. Now, add m

columns $(0,0,0)^T$, $(1,1,1)^T$, $\ldots$, $(m-1,m-1,m-1)^T$ to array $\mathcal{A}$. Array $\mathcal{A}$ now has m^2 columns and clearly, adding these last m columns does not cause the array $\mathcal{A}$ to violent the property of Lemma 8.1.3. Hence the array $\mathcal{A}$ is an $\mathsf{SHF}(3; m^2, m, \{1,2\})$. □

Remark 8.1.6. *The separating hash families obtained from Steiner triple systems of order $m > 3$ are non-isomorphic from those constructed in Construction 8.1.2.*

In the following, a construction of an SHF of type $\{1,2\}$ with $N = 4$ is given.

Construction 8.1.7. [32] The following array is an optimal $2-\mathsf{SHF}(4; m^2, m, \{1,2\})$ for each positive integer $m \geq 2$. i.e., for every two disjoint sets of columns C_1 and C_2 with $|\mathsf{C}_1| = 1$ and $|\mathsf{C}_2| = 2$ there are at least two rows separating C_1 from C_2.

$$\left(\begin{array}{ccccc|ccccccc|c|ccccc}
1 & 1 & \ldots & 1 & 1 & 2 & 2 & 2 & \ldots & 2 & 2 & \ldots & m & m & \ldots & m & m \\
1 & 2 & \ldots & m-1 & m & 1 & 2 & 3 & \ldots & m-1 & m & \ldots & 1 & 2 & \ldots & m-1 & m \\
2 & 3 & \ldots & m & 1 & 3 & 4 & 5 & \ldots & 1 & 2 & \ldots & 1 & 2 & \ldots & m-1 & m \\
m & m-1 & \ldots & 2 & 1 & 1 & m & m-1 & \ldots & 3 & 2 & \ldots & m-1 & m-2 & \ldots & 1 & m
\end{array}\right)$$

The above array is constructed in the following way:

Let $X = \mathbb{Z}_m$ and index the columns of the array $\mathcal{A} = (a_{i,j}) = (\phi_i(x,y))$ with elements of the set $Y = X \times X$. Then define the elements of $\mathcal{A}$ as follows:

$$\begin{aligned}
\phi_1(x,y) &= x \\
\phi_2(x,y) &= y \\
\phi_3(x,y) &= x + y \pmod m \\
\phi_4(x,y) &= x - y \pmod m
\end{aligned}$$

Here is a recursive construction for an SHF of type $\{w_1, w_2\}$.

Construction 8.1.8. [44] Suppose there exists an $\mathsf{SHF}(N_0; n_0, m, \{w_1, w_2\})$ where $gcd\ (n_0, (w_1 w_2)!) = 1$. Then there exists an $\mathsf{SHF}((w_1 w_2 + 1)^j N_0; n_0^{2j}, m, \{w_1, w_2\})$ for any integer $j \geq 0$ which is constructed in the following way:

[4] Let $\mathcal{A}$ be an $\mathsf{SHF}(N_0; n_0, m, \{w_1, w_2\})$. Define $D = (d_{i,j})$ by the rule:

$$d_{i,j} = ij \ mod\ n_0,$$

$0 \leq i \leq \binom{w}{2}$, $0 \leq j \leq n_0 - 1$. For $0 \leq j \leq n_0 - 1$, let A^j denote the array obtained from $\mathcal{A}$ by letting the permutation σ^j act on the columns of $\mathcal{A}$, where $\sigma(i) = i - 1 \ mod\ n_0$.

Now, construct the array $\mathcal{B}$ as follows:

$$\mathcal{B} = \begin{array}{|c|}\hline B_0 \\ \hline B_1 \\ \hline \vdots \\ \hline B_d \\ \hline \end{array} = \begin{array}{|c|c|c|c|}\hline B_{0,0} & B_{0,1} & \cdots & B_{0,n_0-1} \\ \hline B_{1,0} & B_{1,1} & \cdots & B_{1,n_0-1} \\ \hline \vdots & \vdots & \vdots & \vdots \\ \hline B_{d,0} & B_{d,1} & \cdots & B_{d,n_0-1} \\ \hline \end{array}$$

where

$$B_{i,j} = A^{d_{i,j}},$$

$0 \leq i \leq \binom{w}{2}$, $0 \leq j \leq n_0 - 1$. Then $\mathcal{B}$ is an $\mathsf{SHF}((\binom{w}{2}+1)N_0; n_0^2, m, \{w_1, w_2\})$. This procedure can be iterated j times, for any integer $j \geq 2$ and results in an $\mathsf{SHF}((\binom{w}{2}+1)^j N_0; n_0^{2j}, m, \{w_1, w_2\})$.

8.2 New recursive constructions

In this section, we represent two new recursive constructions for separating hash families. In the first construction, we use an existing 2-separating hash family (Definition 8.2.1) on 2 elements and of type $\{2, 2\}$ and construct a new 2-separating hash family.

The idea of the second construction was introduced by Martirosyan and Tran Van Trung in [36] for perfect hash families. We now use it to construct a separating hash family with new parameters from an existing separating hash family.

8.2.1 First construction

Li et. al [32] introduced the concept of w-separating hash families and used these objects in constructing cover-free families.

Definition 8.2.1. Let n, m, w, w_1 and w_2 be positive integers and X and Y be arbitrary sets with $|X| = n$ and $|Y| = m$. An $(n, m, \{w_1, w_2\}) - w-$separating hash family is a set of functions $\mathcal{F}$, such that $f : X \to Y$ for each $f \in \mathcal{F}$ and for any $\mathsf{C}_1, \mathsf{C}_2 \subseteq X$ with $|\mathsf{C}_1| = w_1$ and $|\mathsf{C}_2| = w_2$ and $\mathsf{C}_1 \cap \mathsf{C}_2 = \emptyset$, there exists at least w functions $f \in \mathcal{F}$ such that

$$\{f(y) : y \in \mathsf{C}_1\} \cap \{f(y) : y \in \mathsf{C}_2\} = \emptyset.$$

To denote an $(n, m, \{w_1, w_2\}) - w-$separating hash family with $|\mathcal{F}| = N$, we use the notation $w - \mathsf{SHF}(N; n, m, \{w_1, w_2\})$.
Like separating hash families, a $w - \mathsf{SHF}(N; n, m, \{w_1, w_2\})$ can be depicted as an $N \times n$ matrix with entries from $\{1, \ldots, m\}$ having the following property: For any two distinct sets C_1 and C_2 of w_1 and w_2 columns (respectively), there exist at least w rows such that the entries in the columns of C_1 are distinct from the entries in the columns of C_2.
Assume that $\mathcal{F}$ is a 2-separating hash family of type $\{2, 2\}$ with N rows, n columns and 2 symbols. Let $M = \{0, 1\}$ be the symbol set of $\mathcal{F}$. We use the notation $\overline{a}$ to show the complement of the element $a = 0, 1$ which is defined as follows:

$$\overline{0} = 1 \ , \ \overline{1} = 0.$$

In the following construction, the matrix $\overline{\mathcal{A}}$ is defined as follows:

If $\mathcal{A} = (a_{i,j})$ is an $N \times n$ matrix with element set $\{0, 1\}$ then $\overline{\mathcal{A}} = (\overline{a_{i,j}})_{N \times n}$.
In the following construction, we consider that a $2 - \mathsf{SHF}(N; n, 2, \{2, 2\})$ exists and use it to construct a $2 - \mathsf{SHF}(2N + 2; 2n, 2, \{2, 2\}$.

Construction 8.2.2. Assume that $\mathcal{A}$ is the matrix representation of a $2-\mathsf{SHF}(N; n, 2, \{2,2\})$. The following recursive construction is a $2-\mathsf{SHF}(2N+2; 2n, 2, \{2,2\})$:

$$\mathcal{B} = \left(\begin{array}{cccc|cccc} & & \mathcal{A} & & & & \overline{\mathcal{A}} & \\ & & \mathcal{A} & & & & \mathcal{A} & \\ 0 & 0 & \ldots & 0 & 1 & 1 & \ldots & 1 \\ 1 & 1 & \ldots & 1 & 0 & 0 & \ldots & 0 \end{array}\right).$$

Proof. Assume that $\mathsf{C}_1 = \{c_1, d_1\}$ and $\mathsf{C}_2 = \{c_2, d_2\}$ are two disjoint sets of columns. Let $\mathcal{B}_1, \mathcal{B}_2, \mathcal{B}_3$ and $\mathcal{B}_4$ denote the different parts of $\mathcal{A}$ as shown below:

$$\mathcal{B} = \left(\begin{array}{cccc|cccc} & & \mathcal{B}_1 = \mathcal{A} & & & & \mathcal{B}_2 = \overline{\mathcal{A}} & \\ & & \mathcal{B}_3 = \mathcal{A} & & & & \mathcal{B}_4 = \mathcal{A} & \\ 0 & 0 & \ldots & 0 & 1 & 1 & \ldots & 1 \\ 1 & 1 & \ldots & 1 & 0 & 0 & \ldots & 0 \end{array}\right).$$

We also use $\mathcal{B}_\ell$ and $\mathcal{B}_r$ to denote the left and the right part of $\mathcal{B}$ as shown below:

$$\mathcal{B}_\ell = \left(\begin{array}{cccc} & \mathcal{B}_1 = \mathcal{A} & & \\ & \mathcal{B}_3 = \mathcal{A} & & \\ 0 & 0 & \ldots & 0 \\ 1 & 1 & \ldots & 1 \end{array}\right) \text{ and } \mathcal{B}_r = \left(\begin{array}{cccc} & \mathcal{B}_2 = \overline{\mathcal{A}} & & \\ & \mathcal{B}_4 = \mathcal{A} & & \\ 1 & 1 & \ldots & 1 \\ 0 & 0 & \ldots & 0 \end{array}\right).$$

Assume that column c belongs to $\mathcal{B}_\ell$ and column d belongs to $\mathcal{B}_r$. In this proof, we say that c corresponds to d when $c = d \bmod n$.

One of the following cases can happen:

(i) $\mathsf{C}_1, \mathsf{C}_2 \subseteq \mathcal{B}_\ell$: As $\mathcal{A}$ is a 2-separating hash family, there are two rows in $\mathcal{B}_1$ and two rows in $\mathcal{B}_3$ that separate C_1 from C_2.

$\mathsf{C}_1, \mathsf{C}_2 \subseteq \mathcal{B}_r$: 2-separating property of $\mathcal{A}$ results in the existence of two rows in $\mathcal{B}_2$ and two rows in $\mathcal{B}_4$ that separate C_1 from C_2.

(ii) $\mathsf{C}_1 \subseteq \mathcal{B}_\ell$, $\mathsf{C}_2 \subseteq \mathcal{B}_r$ (or $\mathsf{C}_2 \subseteq \mathcal{B}_\ell$, $\mathsf{C}_1 \subseteq \mathcal{B}_r$): The last two rows separate C_1 from C_2.

(iii) $c_1 \in \mathcal{B}_\ell$, $d_1, c_2, d_2 \in \mathcal{B}_r$ (A similar argument can be used when one column belongs to $\mathcal{B}_r$ and three columns belong to $\mathcal{B}_\ell$). The numbers c_1, d_1, c_2 and d_2 satisfy in one of the following conditions:

(iii-a) d_1, c_2 and $d_2 \neq c_1 \bmod n$. Then c_1 corresponds to some column c in $\mathcal{B}_r$ different from d_1, c_2, and d_2. Using the separating property of $\mathcal{A}$, we conclude that there are two rows, namely i, j in $\mathcal{B}_4$ that separate $\{c, d_1\}$ from $\{c_2, d_2\}$. i, j separate C_1 from C_2 also in $\mathcal{B}$.

(iii-b) $d_1 = c_1 \bmod n$. As $\mathcal{A}$ is a separating hash family of type $\{2,2\}$, using Theorem 2.2.6, $\mathcal{A}$ is also a separating hash family of type $\{1,2\}$. Hence,

there exist two rows, $i,\ j$ which separate $\{d_1\}$ from $\{c_2, d_2\}$ in $\mathcal{B}_4$. As d_1 and c_1 have the same elements in rows i and j, these two rows separate C_1 from C_2.

(iii-c) $c_2 = c_1 \bmod n$. In this case we consider d_1, c_2 and d_2 in $\mathcal{B}_2$. Using the separating property of $\mathcal{A}$, hence $\overline{\mathcal{A}}$, there must exist at least two rows i, j separating $\{d_1\}$ from $\{c_2, d_2\}$ in $\mathcal{B}_2$. As there are only two symbols, columns c_2 and d_2 have the same symbol in rows i and j, and different from column d_1. As c_2 is the complement of c_1 in $\mathcal{B}_2$, we conclude that the symbols in rows i and j in column d_1 is the same as the symbols in rows i and j and column c_1. Hence, i and j separate C_1 from C_2.

(iv) $c_1, c_2 \in B_\ell$, $d_1, d_2 \in B_r$.

(iv-a) $c_1, c_2 \neq d_1, d_2 \bmod n$. d_1 and d_2 correspond to two columns c and d different from c_1 and c_2 in $\mathcal{B}_\ell$. Using the separating property of $\mathcal{A}$ we conclude that there are at least two rows that separate $\{c_1, c\}$ from $\{c_2, d\}$ in $\mathcal{B}_3$. As c and d_1, also d and d_2, have the same symbols in rows i and j we conclude that i and j separate C_1 from C_2.

(iv-b) $c_1 = d_1 \bmod n$.

(iv-b-1) $c_2 = d_2 \bmod n$. As $\mathcal{A}$ is separating of type $\{2, 2\}$, it is also separating of type $\{1, 1\}$, (according to Theorem 2.2.6). Hence, there exist at least two rows i, j in $\mathcal{B}_3$ separating c_1 from c_2. As d_1 has the same symbols as c_1 and d_2 has the same elements as c_2 in rows i and j, we conclude that i and j separate $\{c_1, d_1\}$ from $\{c_2, d_2\}$.

(iv-b-2) $c_2 \neq d_2 \bmod n$. d_2 corresponds to a column d in $\mathcal{B}_\ell$ which is different from c_1 and c_2. Using separating property of $\mathcal{A}$, we conclude that there exist two rows i and j separating $\{c_1\}$ from $\{c_2, d_1\}$. As in rows i and j, c_1 and d_1 have the same symbols and d has the same elements as d_2 , these rows separate C_1 from C_2.

(iv-c) $c_1 = d_2 \bmod n$.

(iv-c-1) $c_2 = d_1 \bmod n$. There exist two rows in $\mathcal{B}_1$ separating c_1 from c_2. As c_1 contains the complements of the elements in d_2 and c_2 is the complement of d_1 in rows i and j, we have separation between C_1 and C_2 in these two rows.

(iv-c-2) $c_2 \neq d_1 \bmod n$. We consider the column corresponding to c_2, namely c in $\mathcal{B}_r$. There are two rows, i and j, separating $\{c, d_1\}$ from $\{d_2\}$. It means that in rows i and j columns c and d_1 have the same symbol, a and column d_2 has symbol $\overline{a}$. Then, column c_2 has $\overline{a}$ in these two rows. c_1 which is the complement of d_2 in rows i and j includes a. Hence, $\{c_1, d_1\}$ is separated from $\{c_2, d_2\}$ in rows i and j. □

8.2.2 Second construction

In the following recursive construction, we use an existing separating hash family with parameters $(N; n, m, \{w_1, w_2, \ldots, w_t\})$ to construct an $\mathsf{SHF}(2N, 2n, m+n, \{w_1+1, w_2+1, \ldots, w_t+1\})$.

Construction 8.2.3. Assume that $\mathcal{A}$ is the matrix representation of an $\mathsf{SHF}(N; n, m, \{w_1, w_2, \ldots, w_t\})$ with symbol set V. Let W be a set of symbols with $|W| = n$ such that $W \cap V = \emptyset$. Suppose $\mathcal{B}$ is an $N \times n$ matrix in which each row is a copy of W. We prove that the following matrix is an $\mathsf{SHF}(2N; 2n, m+n, \{w_1+1, w_2+1, \ldots, w_t+1\})$ in which the symbol set is $W \cup V$:

$$\mathcal{D} = \begin{pmatrix} \mathcal{A} & \mathcal{B} \\ \mathcal{B} & \mathcal{A} \end{pmatrix}.$$

Proof. Assume that $\mathsf{C}_1, \mathsf{C}_2, \ldots, \mathsf{C}_t$ are t disjoint sets of columns of $\mathcal{D}$ with $|\mathsf{C}_i| = w_i+1$ for $i = 1, \ldots, t$. Let $\mathcal{D}_\ell$ and $\mathcal{D}_r$ denote the left and the right part of $\mathcal{D}$ as shown below:

$$\mathcal{D}_\ell = \begin{pmatrix} \mathcal{A} \\ \mathcal{B} \end{pmatrix} \text{ and } \mathcal{D}_r = \begin{pmatrix} \mathcal{B} \\ \mathcal{A} \end{pmatrix}.$$

The following cases can happen:

(i) $\mathsf{C}_1, \ldots, \mathsf{C}_t \subseteq \mathcal{D}_\ell$ or $\mathsf{C}_1, \ldots, \mathsf{C}_t \subseteq \mathcal{D}_r$:

If $\mathsf{C}_1, \ldots, \mathsf{C}_t \subseteq \mathcal{D}_\ell$, all the rows $i = N+1, \ldots, 2N$ separate $\mathsf{C}_1, \ldots, \mathsf{C}_t$.
If $\mathsf{C}_1, \ldots, \mathsf{C}_t \subseteq \mathcal{D}_r$, all the rows $i = 1, \ldots, N$ separate $\mathsf{C}_1, \ldots, \mathsf{C}_t$.

(ii) $\mathsf{C}_1, \ldots, \mathsf{C}_s \subseteq \mathcal{D}_l$ and $\mathsf{C}_{s+1}, \ldots, \mathsf{C}_t \subseteq \mathcal{D}_r$:

Without loss of generality, assume that $s \leq t - s$. (In the opposite case we consider the last N rows of the matrix.) Using Theorem 2.2.7 on $\mathcal{A}$, we conclude that $\mathcal{A}$ is an $\mathsf{SHF}(N; n, m, \{w_1+1, w_2+1, \ldots, w_s+1\})$. Hence, there exists some row i, $1 \leq i \leq N$, which separates $\mathsf{C}_1, \ldots, \mathsf{C}_s$ in $\mathcal{A}$. In each row of $\mathcal{B}$ there are n different symbols, i.e. each row of $\mathcal{B}$ separates all sets of columns in $\mathcal{B}$. As the symbol set of $\mathcal{A}$ is different from the symbol set of $\mathcal{B}$ we conclude that row i separates $\mathsf{C}_1, \ldots, \mathsf{C}_t$.

(iii) $\mathsf{C}_1, \ldots, \mathsf{C}_s \subseteq \mathcal{D}_l$, $\mathsf{C}_{s+1}, \ldots, \mathsf{C}_q \subseteq \mathcal{D}_r$ and $\mathsf{C}_{q+1}, \ldots, \mathsf{C}_t$ consist of columns from both $\mathcal{D}_l$ and $\mathcal{D}_r$:

We assume that $s \leq q - s$ (when $s > q - s$, we consider the rows $i = N+1, \ldots, 2N$). For each column set C_i with $i = q+1, \ldots, t$, we have $|\mathsf{C}_i \cap \mathcal{D}_l| \leq w_i$. Using theorems 2.2.6 and 2.2.7, we conclude that there exists some row j, $1 \leq j \leq N$, which separates $\mathsf{C}_1, \mathsf{C}_2, \ldots, \mathsf{C}_s, \mathsf{C}_{q+1} \cap \mathcal{D}_l, \ldots, \mathsf{C}_t \cap \mathcal{D}_l$. As the columns belonging to $\mathcal{D}_r$ have unique symbols in each row, they are separated from all other columns in each row $i = 1, \ldots, N$. Hence, row j separates $\mathsf{C}_1, \mathsf{C}_2, \ldots, \mathsf{C}_t$. □

8.3 Optimal recursive constructions

In this section, we investigate two constructions of separating hash families with parameters $\mathsf{SHF}(\binom{2w-1}{w}; 2w, 2, \{w, w\})$ and $\mathsf{SHF}(\binom{2w-1}{w}; 2w+1, 2, \{w, w\})$ from [44] and prove that they are optimal according to the number of rows.
An $\mathsf{SHF}(\binom{2w-1}{w-1}; 2w, 2, \{w, w\})$ is presented in [44] and it is mentioned that this separating hash family has the minimum number of rows. Here, first we give the construction and then we prove that in any $\mathsf{SHF}(N; 2w, 2, \{w, w\})$ we have $N \geq \binom{2w-1}{w}$.

Theorem 8.3.1. *[44] There exists an* $\mathsf{SHF}(\binom{2w-1}{w-1}; 2w, 2, \{w, w\})$ *where* $w \geq 2$ *is an integer.*

Proof. We construct $\mathcal{A} = (a_{i,j})$, the matrix representation of the separating hash family, in the following way: Let $S_1, \ldots, S_v$ denote the w-subsets $S \subseteq \{1, \ldots, 2w\}$ such that $1 \in S$, implying that $v = \binom{2w-1}{w-1}$. We index the rows of $\mathcal{A}$ by $S_1, \ldots, S_v$ and index the columns by the elements in $\{1, \ldots, 2w\}$. Now the entry in row i and column j of $\mathcal{A}$ is defined as:

$$a_{i,j} = \begin{cases} 1 & \text{if} \quad j \in S_i \\ 0 & \text{if} \quad j \notin S_i \end{cases}$$

We show that $\mathcal{A}$ is an $\mathsf{SHF}(\binom{2w-1}{w-1}; 2w, 2, \{w, w\})$. Let $\mathcal{C}$ denote the set of columns of $\mathcal{A}$ and C_1 and C_2 be two disjoint subsets of $\mathcal{C}$ with $|\mathsf{C}_1| = |\mathsf{C}_2| = w$. As $|\mathcal{C}| = 2w$ we conclude that $\mathsf{C}_1 = \mathcal{C} \setminus \mathsf{C}_2$. Assume that $\mathsf{C}_1 = \{1, c_2, \ldots, c_w\}$. (As $\mathsf{C}_1 \cup \mathsf{C}_2 = \mathcal{C}$, 1 belongs to either C_1 or C_2.) $\{c_2, \ldots, c_w\}$ is a $(w-1)$-subset of $\{2, \ldots, 2w\}$ which means that $\{c_2, \ldots, c_w\} = S_i$ for some $i = 1, \ldots, \binom{2w-1}{w-1}$. Hence row i separates C_1 from C_2. □

We prove in the following theorem that the separating hash family constructed in the proof of Theorem 8.3.1 is optimal regarding the number of rows.

Theorem 8.3.2. *If there exists an* $\mathsf{SHF}(N; 2w, 2, \{w, w\})$ *then* $N \geq \binom{2w-1}{w}$.

Proof. Let $\mathcal{A} = (a_{i,j})$ be an $N \times n$ matrix representing an $\mathsf{SHF}(N; 2w, 2, \{w, w\})$. Row i of $\mathcal{A}$ separates two disjoint sets of columns C_1 and C_2 with $|\mathsf{C}_1| = |\mathsf{C}_2| = w$ if and only if it has entry 0 (or 1) in w columns in C_1 and has entry 1 (or 0) in w columns belonging to C_2. Each row satisfying this condition separates only one pair of disjoint sets of columns C_1 and C_2 with $|\mathsf{C}_1| = |\mathsf{C}_2| = w$. As the number of pairs of disjoint sets C_1 and C_2 with $|\mathsf{C}_1| = |\mathsf{C}_2| = w$ is $\frac{1}{2}\binom{2w}{w}$ we conclude that $N \geq \frac{1}{2}\binom{2w}{w}$ and:

$$\begin{aligned} N &\geq \tfrac{1}{2}\tbinom{2w}{w} \\ &= \tfrac{1}{2}\tfrac{(2w)!}{w!w!} \\ &= \tfrac{1}{2}\tfrac{2w(2w-1)!}{w(w-1)!w!} \\ &= \tbinom{2w-1}{w-1}. \end{aligned}$$

□

In a similar way, we prove the following results:

Theorem 8.3.3. *For any integer $w \geq 2$, there is an* $\mathsf{SHF}(2\binom{2w-1}{w-1}; 2w+1, m, \{w, w\})$.

Proof. Let $\mathcal{A} = (a_{i,j})$ be an $\binom{2w-1}{w-1} \times 2w$ matrix representing an $\mathsf{SHF}(\binom{2w-1}{w-1}; 2w, 2, \{w, w\})$ constructed in the way explained in the proof of Theorem 8.3.1. Then construct a $2\binom{2w-1}{w-1} \times (2w+1)$ matrix as follows:

$$\mathcal{B} = (b_{i,j}) \begin{pmatrix} \mathcal{A} & 0 \\ \mathcal{A} & 1 \end{pmatrix}$$

Now we show that $\mathcal{B}$ is separating of type $\{w, w\}$. Let C_1 and C_2 be two disjoint sets of columns of $\mathcal{B}$ with $|\mathsf{C}_1| = |\mathsf{C}_2| = w$. Let $\mathcal{C}$ denote the set of columns of $\mathcal{A}$. Two cases can happen:

(i) $\mathsf{C}_1, \mathsf{C}_2 \subseteq \mathcal{C}$. According to Theorem 8.3.1, in this case there are two rows in $\mathcal{B}$ which separate C_1 and C_2.

(ii) Column $2w+1 \in \mathsf{C}_1$ (or C_2). From Theorem 8.3.1 and Theorem 2.2.6, we conclude that $\mathcal{A}$ is a separating hash family of type $\{w-1, w\}$. Therefor, there are two rows i and j, $1 \leq i \leq \binom{2w-1}{w-1}$ and $\binom{2w-1}{w-1} + 1 \leq j \leq 2\binom{2w-1}{w-1}$, in $\mathcal{B}$ which separate $\mathsf{C}_1 \setminus \{w+1\}$ from C_2. It is obviously observed that if $b_{i,j} = 1$ for $j \in \mathsf{C}_2$ then row i separates C_1 from C_2 and if $b_{i,j} = 0$ for $j \in \mathsf{C}_2$ then row j separates C_1 from C_2. □

Chapter 9

Future work

In this chapter we mention some possible topics for future research.

Our primary goal of this thesis was to obtain good upper bounds on the number of columns of SHFs. With respect to the general bounds of Chapter 6, there exist examples which show that the exponent cannot be improved in general (for instance Theorem 7.2.6 and Examples 7.3.2, 7.3.3 and 7.3.4). Although in special cases, like Theorem 7.3.5, the exponent is not optimal. However, the constant coefficient in these bounds might be possible to be improved for general type of SHFs. It is interesting to either prove (or construct some examples to show) that the leading constant cannot be improved or to improve it.

A second interesting problem is to find lower bounds on the number of rows of a general SHF. We presented examples of SHFs which have the minimum possible number of rows. However, the problem of obtaining bounds on SHFs has been often considered for upper bounds on the number of columns. Lower bounds on the number of rows of a general SHF have not been studied deeply in literature.

The problem of finding constructions for SHFs is also of interest. It is important to present general constructions for SHFs in which the number of columns is not far from optimality or find optimal SHFs for specified types.

Bibliography

[1] N. Alon, E. Fischer and M. Szegedy, Parent-identifying codes, *Journal of Combinatorial Theory Series A* 95 (2001), 349 – 359.

[2] N. Alon and M. Naor, Rerandomization, witnesses for Boolean matrix multiplication and construction of perfect hash functions, *Algorithmica* 16 (1996), 434 – 449.

[3] N. Alon and U. Stav, New bounds on parent-identifying codes: the case of multiple parents, *Combinatorics, Probability and Computing* 13 (2004), 795 – 807.

[4] M. Atici, S. S. Magliveras, D. R. Stinson and W.-D. Wei, Some recursive constructions for perfect hash families, *Journal of Combinatorial Designs* 4 (1996), 353 – 363.

[5] A. Barg, G. Cohen, S. Encheva, G. Kabatiansky and G. Zemor, A hypergraph approach to the identifying parent property: the case of multiple parents, *SIAM Journal of Discrete Mathematics* 14 (2001), 423 – 431.

[6] A. Barg and G. Kabatiansky, A class of I.P.P. codes with efficient identification, *Journal of Complexity* 20 (2004), 137 – 147.

[7] M. Bazrafshan and Tran van Trung, Bounds for separating hash families, Accepted by *Journal of Combinatorial Theory Series A* (2011).

[8] M. Bazrafshan and Tran van Trung, On optimal bounds for separating hash families, *Preprint IEM* No.4 (2009).

[9] S. G. Barwick, W.-A. Jackson and C. T. Quinn, Optimal linear perfect hash families with small parameters, *Journal of Combinatorial Designs* 12 (2004), 311 – 324.

[10] A. Beimel and Y. Stahl, Robust information-theoretic private information retrieval, *Journal of Crptology* 20 (2007), 295 – 321.

[11] S. R. Blackburn, Combinatorics and threshold cryptography, *Combinatorial Designs and Their Applications* 403 (1999), 49 – 70.

[12] S. R. Blackburn, Perfect hash families: probabilistic methods and explicit constructions, *Journal of Combinatorial Theory Series A* 92 (2000), 54 – 60.

[13] S. R. Blackburn, Frameproof codes, *SIAM Journal of Discrete Mathematics* 16 (2003), 499 – 510.

[14] S. R. Blackburn, An upper bound on the size of a code with the k-identifiable parent property, *Journal of Combinatorial Theory Series A* 102 (2003), 179 – 185.

[15] S. R. Blackburn, A note on separating hash families, *Preprint.* (2007).

[16] S. R. Blackburn, M. Burmester, Y. Desmedt and P. R. Wild, Efficient multiplicative sharing schemes, In EUROCRYPT, (1996), 107 – 118.

[17] S. R. Blackburn, T. Etzion, D. R. Stinson and G. M. Zaverucha, A bound on the size of separating hash families, *Journal of Combinatorial Theory Series A* 115 (2008), 1246 – 1256.

[18] S. R. Blackburn and P. R. Wild, Optimal linear perfect hash families, *Journal of Combinatorial Theory Series A* 83 (1998), 233 – 250.

[19] D. Boneh and J. Shaw, Collusion-secure fingerprinting for digital data, *IEEE Transactions on Information Theory* 44 (1998), 1897 – 1905.

[20] K. A. Bush, Orthogonal arrays of index unity, *Ann. Math. Stat.* 23 (1952) 426–434.

[21] G. Cohen, S. Encheva, S. Litsyn and H. G. Schaathun, Interesting codes and separating codes, *Discret Applies Mathematics* 128 (2003), 75 – 83.

[22] G. Cohen, S. Encheva and H. G. Schaathun, On separating codes, Technical Report 2001D003, TELECOM ParisTech, Ecole Nationale Superieure des Telecommunications, (2001).

[23] G. D. Cohen and H. G. Schaathun, Upper bounds on separating codes. *IEEE Transactions on Information Theory* 50 (2004), 1291 – 1295.

[24] C. J. Colbourn and J. H. Dinitz, editors, *The CRC Handbook of Combinatorial Designs* Chapman and Hall/CRC, Boca Raton, FL, 2nd edition, 2007.

[25] Z. J. Czech, G. Havas and B. S. Majewski, Perfect hashing, *Theoretical Computer Science* 182 (1997), 1 – 143.

[26] J. H. Dinitz, A. Ling and D. R. Stinson, Perfect hash families from transversal designs. *Australasian Journal of Combinatorics* 37 (2007), 233 – 242.

[27] S. Encheva and G. Cohen, Some new p-ary two-secure frameproof codes. *Applied Mathematics Letters* 14 (2001), 177 – 182.

[28] S. Encheva and G. Cohen, Frameproof codes against limited coalitions of pirates, *Theoretical Computer Science* 273 (2002), 295 – 304.

[29] A. Fiat and M. Naor, Broadcast encryption, *Proceedings of CRYPTO 1993, Lecture Notes in Computer Science* 773, 480 – 491, 1993.

[30] A. Fiat and T. Tassa, Dynamic traitor tracing, *Proceedings of CRYPTO 1999, Lecture Notes in Computer Science* 1666, 354–371, 1999.

[31] H. D. L. Hollmann, J. H. van Lint, J-P. Linnartz and L. M. G. M. Tolhuizen, On codes with the identifiable parent property, *Journal of Combinatorial Theory Series A* 82 (1998), 121 – 133.

[32] P. C. Li, R. Wei and G. H. J. van Rees, Constructions of 2-cover-free families and related separating hash families, *Journal of Combinatorial Designs* 14 (2006), 423 – 440.

[33] K. M. Martin, J. Pieprzyk, R. Safavi-Naini, H. Wang and P. R. Wild, Threshold MACs, *In Proceedings of ICISC 2002* volume 2587 of LNCS (2003), 237 – 252.

[34] K. M. Martin, R. Safavi-Naini, H. Wang and P. R. Wild, Distributing the encryption and dycryption of a block cipher, *Designs, Codes and Cryptography* 36 (2005), 263 – 287.

[35] S. Martirosyan, Perfect hash families, identifiable parent property codes and covering arrays, PhD thesis, University of Duisburg-Essen, 2003.

[36] S. Martirosyan and Tran van Trung, Explicit constructions for perfect hash families, *Designs, Codes and Cryptography* 46 (2008), 97 – 112.

[37] K. Mehlhorn, On the program size of perfect and universal hash functions, *In Proceedings of the 23rd Annual Symposium on Foundations of Computer Science* (1982), 170 – 175.

[38] K. Mehlhorn, *Data Structures and Algorithms 1: Sorting and Searching* Springer, 1984.

[39] I. Newman and A. Wigderson, Lower bounds on formula size of Boolean functions using hypergraph entropy, *SIAM Journal of Discrete Mathematics* 8 (1995), 536 – 542.

[40] P. Sarkar and D. R. Stinson, Frameproof and IPP codes, *Lecture Notes in Computer Science* 2247 (2001), 117 – 126 (INDOCRYPT 2001 Proceedings).

[41] H. G. Schathun and G. D. Cohen, A trellis-based bound on (2,1)-separating codes, *Lecture Notes in Computer Science* 3796 (2005), 59 – 67 (Cryptography and Coding 2005).

[42] J. N. Staddon, D. R. Stinson and R. Wei. Combinatorial properties of frameproof and traceability codes, *IEEE Transactions on Information Theory* 47 (2001), 1042 – 1049.

[43] D. R. Stinson, On some methods for unconditionally secure key distribution and broadcast encryption, *Designs, Codes and Cryptography* 12 (1997), 215 – 243.

[44] D. R. Stinson, Tran van Trung and R. Wei, Secure frameproof codes, key distribution patterns, group testing algorithms and related structures, *Journal of Statistical Planning and Inference* 86 (2000), 595 – 617.

[45] D. R. Stinson, R. Wei and K. Chen, On generalized separating hash families, *Journal of Combinatorial Theory Series A* 115 (2008), 105 – 120.

[46] D. R. Stinson, R. Wei and L. Zhu, New constructions for perfect hash families and related structures using combinatorial designs and codes, *Journal of Combinatorial Designs* 8 (2000), 189 – 200.

[47] D. R. Stinson and G. M. Zaverucha, Some improved bounds for secure frameproof codes and related separating hash families, *IEEE Transactions on Information Theory* 54 (2008), 2508 – 2514.

[48] D. R. Stinson and G. M. Zaverucha, New bounds for generalized separating hash families, *Preprint.* (2007) Available onine: http://www.cacr.math.uwaterloo.ca/techreports/2007/cacr2007-21.pdf

[49] D. Tonien and R. Safavi-Naini, Recursive constructions of secure codes and hash families using difference function families, *Journal of Combinatorial Theory Series A* 113 (2006), 664 - 674.

[50] Tran van Trung and S. Martirosyan, New constructions for IPP codes, *Designs, Codes and Cryptography* 35 (2005), 227 – 239.

[51] H. Wang and C. Xing, Explicit constructions of perfect hash families from algebraic curves over finite fields, *Journal of Combinatorial Theory Series A* 93 (2001), 112 - 124.

[52] C. Xing, Asymptotic bounds on frameproof codes, *IEEE Transactions on Information Theory* 48 (2002), 2991 – 2995.

List of Symbols and Abbreviations

Symbol	Description
SHF	separating hash family
PHF	perfect hash family
HF	hash family
FP	frameproof
SFP	secure frameproof
IPP	identifiable parent peoperty
$\mathcal{A}$	matrix representation of an SHF
$\mathcal{A}(r, a)$	the element in row r and column a of $\mathcal{A}$
N	number of rows of an SHF
n	number of columns of an SHF
m	number of symbols of an SHF
$\{w_1, \ldots, w_t\}$	type of an SHF
f	hash function
X	domain of f
Y	range of f
$\|X\|$	cardinality of X
$\mathcal{F}$	family of hash functions
$\mathcal{C}$	the set of columns of $\mathcal{A}$, a code
$\mathcal{Q}$	the set of symbols of a code
$\mathrm{desc}(\mathcal{C}_0)$	the set of descendants of $\mathcal{C}_0$
$\mathrm{desc}_w(\mathcal{C})$	the w-descendant code of $\mathcal{C}$
OA	orthogonal array
(N, t)-staircase	a staircase with parameters N and t
$\mathcal{G}(\mathcal{A})$	a graph defined on the set of columns of an SHF
MOLS	mutually orthogonal Latin squares